EXHIBIT 5
(PART 1 OF 4)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 August 2003 (21.08.2003)

PCT

(10) International Publication Number
WO 03/068054 A2

(51) International Patent Classification[7]: A61B

(21) International Application Number: PCT/US03/04688

(22) International Filing Date: 13 February 2003 (13.02.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/357,031 13 February 2002 (13.02.2002) US

(71) Applicants *(for all designated States except US)*: **THE GOVERNMENT OF THE UNITED STATES OF AMERICA as represented by THE SECRETARY, DEPARTMENT OF HEALTH SERVICES** [US/US]; National Institutes of Health, Office of Technology Transfer, Suite 325, 6011 Executive Boulevard, Rockville, MD 20852-3804 (US). **THE SLOAN KETTERING INSTITUTE FOR CANCER RESEARCH** [US/US]; 1275 York Avenue, New York, NY 10021 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **JAZAERI, Amir, A.** [US/US]; 823 Pizer Street, Houston, TX 77009 (US). **BOYD, Jeffrey** [US/US]; 41 Beechdale Road, Dobbs Ferry, NY 10522 (US). **LIU, Edison, T.** [US/US]; 12 Cuscaden Walk #23-01, Singapore 249694 (US).

(74) Agent: **NOONAN, William, D.**; Klarquist Sparkman, LLP, One World Trade Center, Suite 1600, 121 SW Salmon Street, Portland, OR 97204 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *without international search report and to be republished upon receipt of that report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/068054 A2

(54) Title: IDENTIFICATION OF OVARIAN CANCER TUMOR MARKERS AND THERAPEUTIC TARGETS

(57) Abstract: The present disclosure provides methods for classifying ovarian tumors into *BRCA1-type, BRCA2-type* or non-BRCA-type tumor types by measuring expression levels of a plurality of disclosed ovarian tumor markers. The markers disclosed herein are useful in the diagnosis, staging, detection, and/or treatment of ovarian cancer. Also provided are methods of selecting a treatment regimen by selecting the tumor type. Ovarian cancer-linked logarithmic expression ratios and kits for diagnosis, staging, and detection of ovarian cancer using are also provided.

IDENTIFICATION OF OVARIAN CANCER TUMOR MARKERS AND THERAPEUTIC TARGETS

PRIORITY CLAIM

This application claims the benefit of U.S. Provisional Application No. 60/357,031, filed February 13, 2002, which is incorporated by reference in its entirety herein.

FIELD OF THE DISCLOSURE

The present disclosure is related to diagnosing, prognosing, staging, preventing, and treating disease, particularly ovarian cancer.

BACKGROUND

Ovarian cancer has one of the highest mortality rates of all cancers, due in part to the difficulty of diagnosis. Currently, epithelial ovarian cancer is the leading cause of death resulting from gynecological cancer (see Welsh *et al., PNAS* 98: 1176-1181, 2001). Studies indicate that the five-year survival rates for ovarian cancer are as follows: Stage I (93%), Stage II (70%), Stage III (37%), and Stage IV (25%) (see Holschneider & Berek, *Semin. Surg. Oncol.* 19: 3-10, 2000). Thus, there is a particular need for improved methods of early diagnosis, prognosis, and monitoring of ovarian cancer.

Protein and mRNA levels, and changes in these levels, may be associated with specific types of cancer (and cancer progression). Such association is often specific to the type of cancer, meaning that what is overexpressed in one cancer may be under-expressed (or unchanged) in another. Thus, a collection or set of genes/proteins that are differentially regulated in a specific cancer may be indicative and specifically diagnostic of that type of cancer.

Molecular mechanisms involved in the onset and progression of ovarian cancer remain poorly understood. However, some mutations causing ovarian cancer have been identified. Between 5% and 10% of all breast cancers are hereditary. The remaining 90% to 95% are classified as "sporadic," for which no genetic link to development has been identified.

Breast cancer susceptibility genes *BRCA1* (GenBank Accession # U14680) and *BRCA2* (GenBank Accession # U43746) are tumor suppressor genes. Germ-line mutations of *BRCA1* and *BRCA2* are responsible for approximately 5-10% of all epithelial ovarian cancers (see Li and Karlan, *Curr. Oncol. Rep.* 3:27-32, 2001). Of inherited breast cancers, it is believed that inherited mutations in *BRCA1* or *BRCA2* are responsible as many as 70% of all cases.

Those with inherited mutations in *BRCA1* and *BRCA2* have an approximately 63% lifetime risk of developing breast cancer, whereas the general female population has a 12% lifetime risk. The *BRCA1* and *BRCA2* gene mutations are more often identified in breast cancer patients with poor prognostic factors, which are risk factors that predict for poorer treatment outcomes (*e.g.*, estrogen-receptor-negative tumors, higher growth rates, age less than 35 at onset of disease, breast cancer in both breasts). Development of disease in the opposite breast and ovarian cancer also appear to be

more common in breast cancer patients with *BRCA1* or *BRCA2* mutations. Hence, the presence of *BRCA1* or *BRCA2* mutations may indicate a need for more aggressive therapeutic treatments.

The alleles of *BRCA1* and *BRCA2* must be inactivated before tumor development occurs. *BRCA1* and *BRCA2* are believed to take part in a common pathway involved in maintenance of genomic integrity in cells; however, little is known about the specific molecular mechanisms involved in BRCA mutation associated (BRCA-linked) ovarian carcinogenesis. For example, it is not known whether *BRCA1* and *BRCA2* mutations affect common or unique molecular pathways in ovarian cancer, or if these pathways overlap with those involved in the formation of sporadic tumors. Both BRCA proteins have been implicated in important cellular functions, including embryonic development, DNA damage repair, and transcriptional regulation (see Scully and Livingston, *Nature* 408:429-432, 2000; Zheng *et al.*, *Oncogene* 19:6159-6175, 2000; Welcsh *et al.*, *Trends. Genet.* 16:69-74, 2000; and MacLachlan *et al., J. Biol. Chem.* 275:2777-2785, 2000). *BRCA1* and *BRCA2* have each been implicated in defective homologous recombination DNA repair (see Arvanitis *et al.*, International Journal of Molecular Medicine 10:55-63, 2002), and it is believed that each may be a positive regulator of homologous recombination, with *BRCA2* potentially interacting with Rad51, a central homologous recombination effector protein, and *BRCA1* regulating *GADD45*, a DNA damage response gene.

Patients having cervical and endometrial cancer resulting in defects in homologous recombination pathways have been shown to respond favorably to radiotherapy (Arvantis *et al.*). Therefore, patients having ovarian cancer resulting from a defect in *BRCA1* or *BRCA2* may similarly benefit from radiotherapy treatment. Hence, the ability to classify ovarian cancer patients into groups based upon the underlying mutation provides advantages in selecting potential courses of treatment, and in deciding whether to pursue a more aggressive course of treatment.

In sum, there is a need to better understand patterns of gene expression that trigger ovarian cancer, as well as downstream genes that may serve as indicators of ovarian cancer progression or as potential tumor suppressors.

BRIEF SUMMARY OF THE DISCLOSURE

The present disclosure concerns a method of classifying an ovarian tumor as a BRCA1-like or BRCA2-like or non-BRCA-type tumor, by determining a pattern of expression in the ovarian tumor of a plurality of markers listed in Table 1, wherein the pattern of expression in the ovarian tumor is determined relative to a standard ovarian tissue. The pattern of expression of the markers in the ovarian tumor is then compared to the pattern of expression of the same markers in tissue from a known BRCA1-like or BRCA2-like or non-BRCA-type tumor. A similarity of the pattern of expression in the ovarian tumor to a pattern of expression of the comparison tissue of the known BRCA1-like tumor classifies the ovarian tumor as a BRCA1-like tumor; a similarity of the pattern of expression in the ovarian tumor to a pattern of expression of the known BRCA2-like tumor classifies the ovarian tumor as a BRCA2-like tumor; and a similarity of the pattern of expression in the ovarian

tumor to a pattern of expression of the known sporadic tumor classifies the ovarian tumor as a sporadic tumor.

The patterns of expression are determined, for example, by determining a pattern of over-expression or under-expression of the plurality of markers in the ovarian tumor to over-expression or under-expression of the plurality of markers of the comparison tissue. Alternatively, a pattern of both over-expression and under-expression of the plurality of markers in the ovarian tumor is compared to over-expression and under-expression of the plurality of markers in the comparison tissue.

It has also been discovered that ovarian tumors that do not contain a BRCA1 or BRCA2 mutation may be BRCA-1-like or BRCA2-like in that the pattern of expression of the markers is similar to a tumor having a BRCA-1 or BRCA-2 mutation. Hence tumors that would otherwise be considered "non-BRCA-type" can be classified as BRCA-1-like or BRCA-2-like, which can contribute to decisions about treatment and prognosis even in the absence of the mutation.

Standard ovarian tissue serves as a baseline from which patterns of over expression and under expression can be determined. The "standard" ovarian tissue may be, for example, from an immortalized ovarian cell, ovarian tissue from a subject not having ovarian cancer, a subject not predisposed to developing ovarian cancer, or ovarian tissue from a subject from whom the ovarian tumor was obtained at an earlier point in time. It is also possible for the standard tissue to be tumor tissue taken from a patient at an earlier point in time, for example prior to treatment (for example prior to the administration of chemotherapy). However in most instances the "standard" tissue is "normal" non-tumor ovarian tissue, such as an immortalized ovarian cell line, for example an IOSE cell line.

Many different approaches are described in this disclosure for determining the patterns of expression, and assessing similarities. In specific examples, the patterns of expression are patterns of logarithmic expression ratios, hierarchical clustering patterns, or multidimensional scaling patterns. The patterns may be compared visually or statistically to arrive at conclusions regarding similarity of the patterns. For example, when a multi-dimensional scaling pattern is used to generate a three-dimensional representation of data clusters associated with BRCA1-like, BRCA2-like or non-BRCA-like tumors, the position of a data point obtained from the tumor specimen that is being analyzed can indicate whether the tumor specimen has a pattern of expression associated with one of these groups. If the data point from the tumor specimen is present within or closely associated with one of these clusters, it is assigned a classification the same as the cluster in which is it contained or with which it is associated.

Another approach to comparing patterns of over expression and under expression is to assign different color intensities to standard normal deviation values of the logarithmic expression ratios. Similarities of color patterns can then be used to arrive at a qualitative assignment of a tumor specimen to a classification. In another approach, the logarithmic expression ratios of the plurality markers is compared using compound covariate predictor analysis.

In particular examples discussed herein, a *BRCA1*-like ovarian tumor is differentiated from a non-BRCA-like ovarian tumor by comparing relative logarithmic expression ratios of at least one

marker shown in Table 6. In more particular embodiments, the pattern of expression of all the markers in Table 6 (*CD72*, *SLC25A11*, *LCN2*, *PSTPIP1*, *SIAHBP1*, *UBE1*, *WAS*, *IDH2*, and *PCTK1*) is compared to the pattern of expression of these same markers in the specimen undergoing classification.

In another example, a *BRCA2*-like ovarian tumor is distinguished from a non-BRCA-like ovarian tumor by comparing relative logarithmic expression ratios of at least one marker shown in Table 7, and in some embodiments both of the markers (*LOC51760* and *LRPAP1*). In yet other examples *BRCA1*- and *BRCA2*-like ovarian tumors are distinguished from non-*BRCA*-like ovarian tumors by comparing relative logarithmic expression ratios of at least one marker shown in Table 8, for example *PSTPIP1*, *IDH2*, and *PCTK1*, or all the markers in Table 8. In other examples, a *BRCA1*-like ovarian tumor is distinguished from a *BRCA2*-like ovarian tumor by comparing relative logarithmic expression ratios of at least one marker shown in Table 10, more than one marker shown in Table 10, or all the markers in Table 10.

The disclosed methods also include selecting a treatment strategy based on classifying the ovarian tumor as *BRCA1*-like, *BRCA2*-like or non-*BRCA*-like. For example, the treatment strategy may include selecting a more aggressive treatment regimen for a *BRCA1*-like or *BRCA2*-like tumor (even if the tumor does not contain a *BRCA1* or *BRCA2* mutation). Such treatment regimens can include chemotherapy, radiotherapy, or surgical removal of the tumor and/or surrounding tissue.

In yet other disclosed examples, the expression patterns of a tumor specimen and known comparison tissue are compared using a database of patterns (for example a database of logarithmic expression patterns) associated with *BRCA1*-like, *BRCA2*-like or non-*BRCA*-like ovarian tumors. The database can contain, for example, expression ratios of the plurality of markers in standard tissue. The patterns of the expression ratios of the plurality of markers of the tumor specimen can then be compared to the pattern of expression ratios of the same markers in the standard tissue.

In some examples, comparisons may be made just of patterns of over expression, for one or more markers that is over expressed as listed in Table 5. Alternatively, comparisons may be made just of patterns of under expression. The patterns of expression may be obtained by using nucleic acid sequences of the markers to perform nucleic acid hybridization of specific oligonucleotide probes to the nucleic acid sequences. The markers may be amplified prior to performing nucleic acid hybridization, and expression quantitated to detect a level of differential expression. The markers are conveniently provided on an array, such as a cDNA microarray. In one example the cDNA microarray contains at least 50, 100, 200, 400 or more of the markers listed in Table 1.

The results of these comparisons can be used to diagnose or provide a prognosis of progression of ovarian cancer in a subject. The patterns of expression can also be used to screen for therapeutic agents for the treatment of ovarian cancer, or monitoring response to therapy in a subject, by looking for a return of the patterns of expression of the ovarian tumor toward a non-tumor tissue pattern. Kits are also provided for performing these analyses, and the kit may include arrays with cDNAs of the markers.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the overall expression differences between *BRCA1*-like, *BRCA2*-like, and non-*BRCA*-like ovarian epithelial cancers. **Figure 1A.** Multidimensional scaling model based on the overall gene expression (6,445 filtered spots, Example 1) in *BRCA1*-linked (solid circles), *BRCA2*-linked (open circles), and sporadic tumors (asterisks). **FIG 1B.** The magnitude of differences in gene expression between various tumor groups as revealed by the number of genes differentially expressed among them using the uniform statistical cutoff $P<0.0001$.

Figure 2 illustrates that *BRCA1*- and *BRCA2*-discriminating genes also segregate sporadic ovarian cancers into two groups (*BRCA1*-like and *BRCA2*-like). **Figure 2A.** Hierarchical clustering of 110 non-redundant genes (see Table 9, Addendum) showing significant differential expression between *BRCA1*-linked and (B1) and *BRCA2*-linked (B2) tumors (modified F-test P<0.0001). The red and green color intensities represent standard normal deviation (Z-score) values from the mean expression level of each gene (represented as black) across sixty-one tumor samples (Example 1). **Figure 2A'** is a duplication of Figure 2A, but is printed in grey tones rather than in color. **Figure 2B.** Hierarchical clustering of sporadic and *BRCA*-linked tumor samples based on the expression pattern of the 110 *BRCA*-discriminating patterns of gene expression. The B-, B2-, and C-labeled samples signify *BRCA1*-linked, *BRCA2*-linked, and sporadic tumors, respectively. **Figure 2C.** Hierarchical clustering of sporadic samples in the absence of BRCA-linked tumors reveals two major clusters corresponding to *BRCA1*-type and *BRCA2*-type patterns of gene expression.

Figure 3 shows molecular profiles of sixty-one tumors as defined by the genes whose expression significantly differentiated *BRCA1* and *BRCA2* tumors ($P<0.0001$) (see Example 1, and Table 9). The red and green color intensities represent expression levels shown as standard normal deviation (Z-score) values from the mean expression level of each gene (represented as black) across sixty-one tumor samples. The genes are numbered consecutively 1-61 in Figure 3A, and 62-116 in Figure 3B. **Figure 3A' and Figure 3B'** are duplications of **Figure 3A** and **Figure 3B**, respectively, but are printed in grey tones rather than in color. **Figure 3C** shows the correlation of the designated rows to genes and SEQ ID NOs for the molecular profile in Figure 3A and **Figure 3D** shows the correlation of the designated rows to genes and SEQ ID NOs for the molecular profile in Figure 3B.

Figure 4 shows gene expression differences between BRCA-linked and sporadic tumors. A modified F-test with a statistical significance level of $P<0.0001$ was used to evaluate genes differentially expressed between tumor types. The red and green color intensities represent expression levels shown as standard normal deviation (Z-score) values from the mean expression level of each gene (represented as black) across all sixty-one tumor samples. Each gene name is followed by the corresponding I.M.A.G.E. clone number spotted on the array. **Figure 4A.** Genes differentially expressed between *BRCA1*-linked (B) and sporadic (C) samples. Genes located on Xp11 appear in red. **Figure 4B.** Examples of genes differentially expressed between *BRCA2*-linked (B2) and sporadic (C) samples. **Figure 4C.** Examples of differentially expressed genes between the combined *BRCA1*- and *BRCA2*-linked group (B and B2, respectively) and the sporadic (C) samples. **FIG 4A-C'** is a duplication of Figure 4A-C, but is printed in grey tones rather than in color.

Figure 4D *BRCA1*-linked tumors exhibit significantly higher expression levels ($P<0.001$) of all six genes mapped to Xp11.23 compared to the sporadic cancers. Error bars reflect standard error.

Figure 5 is a bar graph showing an evaluation of gene expression patterns common to *BRCA*-linked and sporadic tumors. **Figure 5A** shows the expression of twenty-five genes that showed two-fold or greater down-regulation as compared to the IOSE reference cell line. **Figure 5B** shows the expression of twenty-five genes that showed two-fold or greater up-regulation as compared to the IOSE reference cell line. Error bars reflect standard error. (*FOS*, *HE4* and *CD24*) have been previously reported to be overexpressed in ovarian cancers. Several of the overexpressed genes that have been demonstrated to be interferon-responsive are presented in italics. The * symbol denotes immediate-early response genes.

Figure 6 is a series of bar graphs illustrating semi-quantitative RT-PCR (sqRT-PCR) analysis of gene expression confirms the cDNA microarray data. Expression patterns of select genes were examined using sqRT-PCR in representative *BRCA1*-linked (bars 1-5), *BRCA2*-linked (bars 6-10), and sporadic (bars 11-15) samples. The expression level of each gene in the tumor samples was compared to those of normal postmenopausal ovary (N) and the reference IOSE cells (R). All data has been normalized to β-actin is presented as fold expression compared to the IOSE reference RNA. **Figure 6A** shows results for genes HE4, RSG1, and CD74. **Figure 6B** shows results for genes ZFP36, TOP2A and HLA-DRB1.

BRIEF DESCRIPTION OF THE TABLES

TABLE 1 (see Addendum) lists 822 ovarian cancer-related nucleic acid molecules that show altered expression in ovarian cancer. The nucleic acids are identified by their SEQ ID NO, their gene name (if one has been assigned), the I.M.A.G.E Clone ID number associated with the nucleic acid sequence, the UniGene number (if one has been assigned), and a description of the gene (if known). Because more than one GenBank Accession Number is sometimes provided for a given nucleic acid molecule, the Table groups the SEQ ID NO assigned to each GenBank Accession Number with nucleic acid molecule. For example, the entry for *BCKDHB* in Table 1 provides SEQ ID NOs: 16-17 (represented by GenBank Accession number AA427739 and GenBank Accession number AA434304). Each of the 822 SEQ ID NOs are included in the attached sequence listing.

TABLE 2 catalogs the clinicopathologic features of the tumor samples in a study of sixty-one cases of pathologically-confirmed epithelial ovarian adenocarcinoma.

TABLE 3 lists representative gene-specific primer sequences used to amplify RNA for analysis by semi-quantitative PCR.

TABLE 4 (see Addendum) lists markers that were under-expressed in ovarian cancer in a comparison of ovarian epithelial cancer cells to normal postmenopausal ovarian tissue.

TABLE 5 (see Addendum) lists markers that were ove-rexpressed in ovarian cancer in a comparison of ovarian epithelial cancer cells to normal postmenopausal ovarian tissue.

TABLE 6 (see Addendum) lists markers that were differentially expressed between *BRCA1*-linked and sporadic tumors in a comparison to reference immortalized ovarian epithelial cells.

TABLE 7 (see Addendum) lists markers that were differentially expressed between *BRCA2*-linked and sporadic tumors in a comparison to reference immortalized ovarian epithelial cells.

TABLE 8 (see Addendum) lists markers that were differentially expressed between combined *BRCA1*-linked and *BRCA2*-linked versus sporadic tumors in a comparison to reference immortalized ovarian epithelial cells.

TABLE 9 (see Addendum) lists markers that were differentially expressed between *BRCA1*-linked and *BRCA2*-linked tumors in a comparison to reference immortalized ovarian epithelial cells.

TABLE 10 (see Addendum) lists markers that can be used to segregate *BRCA1*-like from *BRCA2*-like tumor types using compound covariate prediction analysis.

TABLE 11 (see Addendum) lists the results of compound covariate predictor analysis for the sixty-one tumors disclosed herein, analyzed using the markers in Table 10.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING

The nucleic and amino acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and single letter code for amino acids, as defined in 37 C.F.R. § 1.822. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood as included by any reference to the displayed strand. In the accompanying sequence listing:

SEQ ID NO: 1 is a 63-nucleotide synthetic primer containing a T7 RNA polymerase binding site.

SEQ ID NOs: 2 and 3 are *ACTB* gene-specific primers used for amplification during semi-quantitative RT-PCR.

SEQ ID NOs: 4 and 5 are *HE4* gene-specific primers used for amplification during semi-quantitative RT-PCR.

SEQ ID NOs: 6 and 7 are *ZFP36* gene specific primers used for amplification during semi-quantitative RT-PCR.

SEQ ID NOs: 8 and 9 are *RGS1* gene specific primers used for amplification during semi-quantitative RT-PCR.

SEQ ID NOs: 10 and 11 are *CD74* gene specific primers used for amplification during semi-quantitative RT-PCR.

SEQ ID NOs: 12 and 13 are *TOP2A* gene specific primers used for amplification during semi-quantitative RT-PCR.

SEQ ID NOs: 14 and 15 are *HLA-DRB1* gene specific primers used for amplification during semi-quantitative RT-PCR.

SEQ ID NOs: 16 through 822 are ovarian cancer-related nucleic acid molecules that show altered expression in ovarian cancer. These nucleic acid molecules are listed in Table 1, and their sequence information is provided in the attached sequence listing.

DETAILED DESCRIPTION

I. Abbreviations

cDNA:	complementary DNA
DNA:	deoxyribonucleic acid
ELISA:	enzyme-linked immunosorbent assay
EST:	expressed sequence tag
I.M.A.G.E.:	Integrated Molecular Analysis of Genomes and their Expression Consortium
IOSE:	immortalized ovarian surface epithelial cell lines
MDS:	multidimensional scaling
PCR:	polymerase chain reaction
RIA:	radioimmunoassay
RNA:	ribonucleic acid
RT-PCR:	reverse transcription-polymerase chain reaction
siRNA:	small inhibitory RNA molecule
sqRT-PCR:	semi-quantitative reverse transcription-polymerase chain reaction
STS:	sequence-tagged site

II. Terms

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found in Benjamin Lewin, *Genes V*, published by Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

In accordance with the present disclosure, conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art are used. Such techniques are fully explained in the literature (see, *e.g.*, Sambrook *et al.*, 1989. *Molecular cloning, a laboratory manual.* 2nd ed. Cold Spring Harbor Laboratory, Cold spring Harbor, New York; Glover, 1985, *DNA Cloning: A practical approach*, volumes I and II oligonucleotide synthesis, MRL Press, LTD., Oxford, U.K.; Hames and Higgins, 1985, *Transcription and translation*; Hames and Higgins, 1984, Animal Cell Culture; Freshney, 1986, *Immobilized Cells And Enzymes*, IRL Press; and Perbal, *A Practical Guide to Molecular Cloning*, John Wiley & Sons, New York, 1988).

In order to facilitate review of the various embodiments of the invention, the following explanations of specific terms are provided:

Altered expression or **differential expression** refers to expression of a nucleic acid (*e.g.*, mRNA or protein) in a subject or biological sample from a subject that deviates from that expression in a subject or biological sample from a subject having normal (wild-type) characteristics for the biological condition associated with the nucleic acid. Normal expression can be found in a control, a standard for a population, etc. For instance, where the altered expression manifests as a diseased condition, such as growth of a tumor or neoplasia or onset of a cancer such as ovarian cancer, characteristics of normal expression might include an individual who is not suffering from the condition (*e.g.*, a subject not displaying neoplasia growth or not having ovarian cancer), a population

standard of individuals believed not to be suffering from the disease, etc. For instance, certain altered expression (such as altered expression of a *BRCA* nucleic acid), can be described as being associated with the biological conditions of altered (*e.g.*, over-expressed or under-expressed) nucleic acid expression and a tendency to develop gynecological cancer, such as ovarian cancer. Likewise, altered expression may be associated with a disease. The term "associated with" includes an increased risk of developing the disease.

Controls or standards (*e.g.*, a reference cell line, such as immortalized epithelial ovarian cells) for comparison to a sample (*e.g.*, an ovarian cancer tumor), for the determination of altered expression, include samples believed to be normal for the studied characteristic, as well as laboratory values, even though possibly arbitrarily set, keeping in mind that such values may vary from laboratory to laboratory. Laboratory standards and values may be set based on a known or determined population value and may be supplied in the format of a graph or table that permits easy comparison of measured, experimentally determined values.

When used in reference to a nucleic acid, **amplification** includes techniques that increase the number of copies of a nucleic acid molecule in a sample or specimen. An example of amplification is the polymerase chain reaction, in which a biological sample collected from a subject is contacted with a pair of oligonucleotide primers, under conditions that allow for the hybridization of the primers to nucleic acid template in the sample. The primers are extended under suitable conditions, dissociated from the template, and then re-annealed, extended, and dissociated to amplify the number of copies of the nucleic acid. The product of *in vitro* amplification can be characterized by electrophoresis, restriction endonuclease cleavage patterns, oligonucleotide hybridization or ligation, and/or nucleic acid sequencing, using standard techniques. Other examples of *in vitro* amplification techniques include strand displacement amplification (*see* U.S. Patent No. 5,744,311); transcription-free isothermal amplification (*see* U.S. Patent No. 6,033,881); repair chain reaction amplification (*see* WO 90/01069); ligase chain reaction amplification (*see* EP-A-320 308); gap filling ligase chain reaction amplification (*see* U.S. Patent No. 5,427,930); coupled ligase detection and PCR (*see* U.S. Patent No. 6,027,889); and NASBA™ RNA transcription-free amplification (*see* U.S. Patent No. 6,025,134).

An **array** is an arrangement of molecules, particularly biological macromolecules (such as polypeptides or nucleic acids) or cell or tissue samples, in addressable locations on or in a substrate. The array may be regular (arranged in uniform rows and columns, for instance) or irregular. The number of addressable locations on the array can vary, for example from a few (such as three) to more than 50, 100, 200, 500, 1000, 10,000, or more. A **microarray** is an array that is miniaturized so as to require or be aided by microscopic examination for evaluation or analysis. A **cDNA microarray** is an array of multiple cDNA molecules, fixed in addressable locations, to which complementary nucleic acids in applied samples may hybridize (see Hegde *et al.*, *Biotechniques* 29(3): 548-562, 2000). cDNA microarrays of the disclosure provide for qualitative and quantitative analysis of gene expression of the molecules contained in the array.

Within an array, each arrayed sample (feature) is addressable, in that its location can be reliably and consistently determined within the at least two dimensions of the array. Thus, in ordered arrays the location of each sample is assigned to the sample at the time when it is applied to the array, and a key may be provided in order to correlate each location with the appropriate target or feature position. Often, ordered arrays are arranged in a symmetrical grid pattern, but samples could be arranged in other patterns (*e.g.*, in radially distributed lines, spiral lines, or ordered clusters). Addressable arrays usually are computer readable, in that a computer can be programmed to correlate a particular address on the array with information about the sample at that position (*e.g.*, expression data, including for instance signal intensity as well as the identity of the sample). In some examples of computer readable formats, the individual features in the array are arranged regularly, for instance in a Cartesian grid pattern, which can be correlated to address information by a computer.

The sample application location on an array (the "feature") may assume many different shapes. Thus, though the term "spot" may be used herein, it refers generally to a localized placement of molecules or tissue or cells, and is not limited to a round or substantially round region. For instance, substantially square regions of application can be used with arrays encompassed herein, as can be regions that are, for example, substantially rectangular, triangular, oval, irregular, or another shape. Within a single array, feature shapes do not usually vary, though they will in some embodiments.

In certain example arrays, one or more features will occur on the array a plurality of times (*e.g.*, twice) to provide internal controls.

A **biological sample** is any sample in which the presence of a protein and/or ongoing expression of a protein may be detected. Suitable biological samples include samples containing genomic DNA or RNA (including mRNA), obtained from body cells of a subject, such as but not limited to those present in peripheral blood, urine, saliva, cells obtained by pap smear, sera, tissue biopsy, surgical specimen, amniocentesis samples and autopsy material.

A ***BRCA1*-like tumor** is a tumor in which the gene expression pattern is substantially similar to the gene expression pattern in a tumor from a subject who has a mutation in *BRCA1*. Similarly, a ***BRCA2*-like tumor** is a tumor in which the gene expression pattern is substantially similar to the gene expression pattern in a tumor from a subject who has a mutation in *BRCA2*. As described herein, sporadic tumors may share gene expression patterns with BRCA-linked and or BRCA2-linked tumors. Hence, sporadic and other tumors (such as tumors for which no BRCA genetic test has been conducted) that have gene expression patterns similar to a BRCA1-linked tumor are "BRCA1-like" tumors.

A **cancer** is a biological condition in which a malignant tumor or other neoplasm has undergone characteristic anaplasia with loss of differentiation, increased rate of growth, invasion of surrounding tissue, and/or which is capable of metastasis.

The term cancer includes ovarian cancer, such as ovarian epithelial cancer, which originates in the ovaries and may manifest as epithelial tumors, germ cell tumors, or stromal tumors. Also included are different stages of a single cancer, for instance both primary and recurrent ovarian

cancer, and cancer at any progressive stage, such as Stages I-IV. Ovarian cancer is considered a gynecological cancer.

A subject may be classified into an **ovarian cancer stage** based upon evaluation of a biological sample from the subject for indices known in the art or disclosed herein as being indicative of that stage of ovarian cancer. For example, a subject may be classified as having a cancer state of cancer-free, active ovarian cancer (*i.e.,* stage I, II, III, or IV ovarian cancer), or in remission from previous ovarian cancer.

cDNA is a piece of DNA lacking internal, non-coding segments (introns) and regulatory sequences that determine transcription. cDNA is generally synthesized in the laboratory by reverse transcription from messenger RNA extracted from cells.

Compound covariate prediction analysis is a method of predicting into which of two groups a sample will be assinged using a given statistical signficance cutoff (*e.g.*, $P<0.0005$). The method creates a multivariate predictor for one of two classes to each sample and includes in the multivariate predictor only those components (*e.g.*, nucleic acids expressing on a cDNA microarray) that meet the statistical signficance cutoff. The multivariate predictor is a weighted linear combination of logarithmic ratios for components that are univariately significant. The weight consists of the univariate t-statistics for comparing the classes.

DNA is a polymer that comprises the genetic material of most living organisms (some viruses have genomes comprising RNA). The repeating units in most natural DNA polymers are four different nucleotides, each of which comprises one of the four bases, adenine, guanine, cytosine and thymine, bound to a deoxyribose sugar to which a phosphate group is attached. Triplets of nucleotides (referred to as codons) code for each amino acid in a polypeptide, or for a stop signal. The term codon is also used for the corresponding (and complementary) sequences of three nucleotides in the mRNA into which the DNA sequence is transcribed.

Unless otherwise specified, any reference to a DNA molecule is intended to include the reverse complement of that DNA molecule. Except where single-strandedness is required by the text herein, DNA molecules, though written to depict only a single strand, encompass both strands of a double-stranded DNA molecule. Thus, a reference to the nucleic acid molecule that encodes a specific protein, or a fragment thereof, encompasses both the sense strand and its reverse complement. Thus, for instance, it is appropriate to generate primers from the reverse complement sequence of the disclosed nucleic acid molecules.

An **expressed sequence tag (EST)** is a unique stretch of DNA within a coding region of a gene that is useful for identifying full-length genes and serves as a landmark for gene mapping. An EST is a sequence tagged site (STS) derived from cDNA.

Expression of a gene is the process by which the coded information of a gene is converted into an operational or non-operational part of a cell, often including the synthesis of a protein. Gene expression can be influenced by external signals. For instance, exposure of a cell to a hormone may stimulate expression of a hormone-induced gene. Different types of cells may respond differently to an identical signal.

Expression of a gene also may be regulated in the pathway from DNA to RNA to protein. Ways in which regulation occurs include through controls acting on transcription, translation, RNA transport and processing, degradation of intermediary molecules such as mRNA, or through activation, inactivation or compartmentalization or degradation of specific protein molecules after they have been made.

Changes in gene expression may be associated with specific types of cancer (and cancer progression). Such association is fairly specific to the type of cancer, and thus what is overexpressed in one cancer may be underexpressed (or unchanged) in another.

The expression of several genes may be grouped into an **expression pattern** or **expression profile**. Such patterns or profiles may be unique to an individual sample depending upon certain factors, for instance biological stimuli introduced into the subject from which the sample was taken (*e.g.*, a hormone) or ongoing disease within the subject (*e.g.*, ovarian cancer). Thus, a collection or set of genes/proteins that are differentially regulated in a specific cancer may be indicative and specifically diagnostic of that type of cancer. In addition, specific expression patterns may indicate particular mutations within the individual that correlate and/or cause the disease, for instance a mutation in *BRCA1* or *BRCA2*, or may indicate a larger class of disease, such as a *BRCA1*-like or *BRCA2*-like cancer. Furthermore, changing the expression patterns of these genes to restore the normal state, or bring the condition closer to the normal state in one or more characteristic, may constitute a treatment for cancer.

As disclosed herein, the expression pattern of an unknown tumor may be compared to the expression pattern of known BRCA1-linked and BRCA2-linked markers to determine if the expression patterns are sufficiently similar to classify the unknown as a BRCA1-like or BRCA2-like tumor.

Gene amplification or **genomic amplification** is an increase in the copy number of a gene or a fragment or region of a gene or associated 5' or 3' region, as compared to the copy number in normal tissue. An example of a genomic amplification is an increase in the copy number of an oncogene. A "gene deletion" is a deletion of one or more nucleic acids normally present in a gene sequence and, in extreme examples, can include deletions of entire genes or even portions of chromosomes.

A **gene expression fingerprint (or profile)** is a distinct or identifiable pattern of gene expression, for instance a pattern of high and low expression of a defined set of genes or gene-indicative nucleic acids such as ESTs; in some instances, as few as one or two genes may provide a profile, but often more genes are used in a profile, for instance at least three, at least 5, at least 10, at least 20, at least 25, or at least 50 or more. Gene expression fingerprints (also referred to as profiles) can be linked to a tissue or cell type, to a particular stage of normal tissue growth or disease progression, or to any other distinct or identifiable condition that influences gene expression in a predictable way. Gene expression fingerprints can include relative as well as absolute expression levels of specific genes, and often are best viewed in the context of a test sample compared to a baseline or control sample fingerprint. By way of example, a gene expression profile may be read on

an array (*e.g.*, a polynucleotide or polypeptide array). Arrays are now well known, and for instance gene expression arrays have been previously described in published PCT application number WO9948916 ("Hypoxia-Inducible Human Genes, Proteins, and Uses Thereof"), incorporated herein by reference in its entirety.

As disclosed herein, the gene expression profile of an unknown tumor may be compared for similarities and differences to the expression profile of a tumor known to express in a BRCA-like manner (*e.g.*, a BRCA1-like or BRCA2-like tumor).

A **genomic target sequence** is a sequence of nucleotides located in a particular region in the human genome that corresponds to one or more specific genetic abnormalities, such as a nucleotide polymorphism, a deletion, or amplification. The target can be for instance a coding sequence; it can also be the non-coding strand that corresponds to a coding sequence.

Gynecological cancers are cancers of the female reproductive system, and include cancers of the uterus (*e.g.*, endometrial carcinoma), cervix (*e.g.*, cervical carcinoma), ovaries (*e.g.*, ovarian carcinoma, serous cystadenocarcinoma, mucinous cystadenocarcinoma, endometrioid tumors, celioblastoma, clear cell carcinoma, unclassified carcinoma, granulosa-thecal cell tumors, Sertoli-Leydig cell tumors, dysgerminoma, malignant teratoma), vulva (*e.g.*, squamous cell carcinoma, intraepithelial carcinoma, adenocarcinoma, fibrosarcoma, melanoma), vagina (*e.g.*, clear cell carcinoma, squamous cell carcinoma, botryoid sarcoma), embryonal rhabdomyosarcoma, and fallopian tubes (*e.g.*, carcinoma).

An **isolated** biological component (such as a nucleic acid, peptide or protein) has been substantially separated, produced apart from, or purified away from other biological components in the cell of the organism in which the component naturally occurs, *i.e.*, other chromosomal and extrachromosomal DNA and RNA, and proteins. Nucleic acids, peptides and proteins that have been isolated thus include nucleic acids and proteins purified by standard purification methods. The term also embraces nucleic acids, peptides and proteins prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids.

A **marker** is a diagnostic indicator of disease. A marker may consist of any signal indicating the presence of the disease, *e.g.*, a physiological change in the body of a subject or increased or decreased levels of a substance such as a protein correlated to the disease. Markers are often found in body fluid samples from a subject. By way of example, prostate specific antigen is a tumor marker used to detect progression of prostate cancer. The molecules disclosed herein, for instance in Table 1 are useful as tumor markers for diagnosing, prognosing, staging, preventing, and treating cancerous disease, such as ovarian cancer.

A **mutation** includes any change of the DNA sequence within a gene or chromosome. In some instances, a mutation will alter a characteristic or trait (phenotype), but this is not always the case. Types of mutations include base substitution point mutations (*e.g.*, transitions or transversions), deletions, and insertions. Missense mutations are those that introduce a different amino acid into the sequence of the encoded protein; nonsense mutations are those that introduce a new stop codon. In the case of insertions or deletions, mutations can be in-frame (not changing the frame of the overall

sequence) or frame shift mutations, which may result in the misreading of a large number of codons (and often leads to abnormal termination of the encoded product due to the presence of a stop codon in the alternative frame).

This term specifically encompasses variations that arise through somatic mutation, for instance those that are found only in disease cells, but not constitutionally, in a given individual. Examples of such somatically-acquired variations include the point mutations that frequently result in altered function of various genes that are involved in development of cancers. This term also encompasses DNA alterations that are present constitutionally, that alter the function of the encoded protein in a readily demonstrable manner, and that can be inherited by the children of an affected individual. In this respect, the term overlaps with "polymorphism," as defined below, but generally refers to the subset of constitutional alterations that have arisen within the past few generations in a kindred and that are not widely disseminated in a population group. In particular embodiments, the term is directed to those constitutional alterations that have major impact on the health of affected individuals, such as those resulting in onset of a disease such as a gynecological cancer.

An **oligonucleotide** is a plurality of joined nucleotides joined by native phosphodiester bonds, between about 6 and about 300 nucleotides in length. An oligonucleotide analog refers to moieties that function similarly to oligonucleotides but have non-naturally occurring portions. For example, oligonucleotide analogs can contain non-naturally occurring portions, such as altered sugar moieties or inter-sugar linkages, such as a phosphorothioate oligodeoxynucleotide. Functional analogs of naturally occurring polynucleotides can bind to RNA or DNA, and include peptide nucleic acid (PNA) molecules.

Particular oligonucleotides and oligonucleotide analogs can include linear sequences up to about 200 nucleotides in length, for example a sequence (such as DNA or RNA) that is at least 6 bases, for example at least 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100 or even 200 bases long, or from about 6 to about 50 bases, for example about 10-25 bases, such as 12, 15 or 20 bases.

A **neoplasm** is a new and abnormal growth, particularly a new growth of tissue or cells in which the growth is uncontrolled and progressive. A tumor is an example of a neoplasm.

A **non-BRCA-type** tumor is a tumor in which the gene expression pattern of the BRCA1-linked and BRCA2-linked markers disclosed in Table 1 is not similar to either a BRCA1-like or BRCA2-like gene expression pattern.

A **nucleic acid** is a deoxyribonucleotide or ribonucleotide polymer in either single or double stranded form, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides.

A **nucleic acid sequence (or polynucleotide)** is a DNA or RNA molecule, and includes polynucleotides encoding full-length proteins and/or fragments of such full length proteins which can function as a therapeutic agent.

Nucleotide includes, but is not limited to, a monomer that includes a base linked to a sugar, such as a pyrimidine, purine or synthetic analogs thereof, or a base linked to an amino acid, as in a

peptide nucleic acid (PNA). A nucleotide is one monomer in a polynucleotide. A nucleotide sequence refers to the sequence of bases in a polynucleotide.

A first nucleic acid sequence is **operably linked** with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein-coding regions, in the same reading frame.

An **ovarian cancer-related molecule** includes nucleic acids (such as DNA or RNA or cDNA) and proteins that are altered (for example by mutation or abnormal expression) in ovarian cancer.

Pharmaceutically acceptable carriers include compositions and formulations suitable for pharmaceutical delivery of the nucleotides and proteins herein disclosed. Martin, *Remington's Pharmaceutical Sciences*, published by Mack Publishing Co., Easton, PA, 19th Edition, 1995, describes conventional pharmaceutically acceptable carriers.

In general, the nature of the carrier will depend on the particular mode of administration being employed. For instance, parenteral formulations usually comprise injectable fluids that include pharmaceutically and physiologically acceptable fluids such as water, physiological saline, balanced salt solutions, aqueous dextrose, glycerol or the like as a vehicle. For solid compositions (*e.g.*, powder, pill, tablet, or capsule forms), conventional non-toxic solid carriers can include, for example, pharmaceutical grades of mannitol, lactose, starch, or magnesium stearate. In addition to biologically-neutral carriers, pharmaceutical compositions to be administered can contain minor amounts of non-toxic auxiliary substances, such as wetting or emulsifying agents, preservatives, and pH buffering agents and the like, for example sodium acetate or sorbitan monolaurate.

Primers are short nucleic acids, preferably DNA oligonucleotides 10 nucleotides or more in length, which are annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, *e.g.*, by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods known in the art.

Primers as used in the present disclosure preferably comprise at least 10 nucleotides of the nucleic acid sequences that are shown to encode specific proteins. In order to enhance specificity, longer primers may also be employed, such as primers that comprise 15, 20, 30, 40, 50, 60, 70, 80, 90 or 100 consecutive nucleotides of the disclosed nucleic acid sequences. Methods for preparing and using probes and primers are described in the references, for example Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York; Ausubel *et al.* (1987) *Current Protocols in Molecular Biology*, Greene Publ. Assoc. & Wiley-Intersciences; Innis *et al.* (1990) *PCR Protocols, A Guide to Methods and Applications*, Innis *et al.* (Eds.), Academic Press, San Diego, CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge, MA).

When referring to a primer, the term *specific for (a target sequence)* indicates that the primer hybridizes under stringent conditions substantially only to the target sequence in a given sample comprising the target sequence.

A **probe** comprises an isolated nucleic acid attached to a detectable label or other reporter molecule. Typical labels include radioactive isotopes, enzyme substrates, co-factors, ligands, chemiluminescent or fluorescent agents, haptens, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed, *e.g.*, Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, CSHL, New York, 1989) and Ausubel *et al.* (In *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1998).

A **protein** is a biological molecule expressed by a gene and comprised of amino acids.

A **purified** molecule is one that has been purified relative to its original environment. The term "purified" does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified protein preparation is one in which the protein referred to is more pure than the protein in its natural environment within a cell or within a production reaction chamber (as appropriate). Non-limiting examples of purified molecules are those that are 50%, 75%, or 90% pure.

A **recombinant nucleic acid** is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, *e.g.*, by genetic engineering techniques such as those described in Sambrook *et al.*, In *Molecular Cloning: A Laboratory Manual*, CSHL, New York, 1989. The term recombinant includes nucleic acids that have been altered solely by deletion of a portion of the nucleic acid. For instance, a plasmid is recombinant if some portion of the naturally occurring plasmid has been deleted. Equally, if the sequence of such a plasmid has been altered, for example by a nucleotide substitution (or addition or deletion), that plasmid is said to be recombinant.

Sequence identity is the similarity between two nucleic acid sequences, or two amino acid sequences is expressed in terms of the similarity between the sequences, otherwise referred to as sequence identity. Sequence identity is frequently measured in terms of percentage identity (or similarity or homology); the higher the percentage, the more similar are the two sequences. Methods of alignment of sequences for comparison are well-known in the art. Various programs and alignment algorithms are described in: Smith and Waterman, *J. Theor. Biol.* 91(2): 379-380, 1981; Needleman and Wunsch, *J. Mol. Bio.* 48:443-453, 1970; Pearson and Lipman, *Methods in Molec. Biology* 24: 307-331, 1988; Higgins and Sharp, *Gene* 73:237-244, 1988; Higgins and Sharp, *CABIOS* 5:151-153, 1989; Corpet *et al.*, *Nucleic Acids Research* 16:10881-10890, 1988; Huang *et al.*, *Computer Applications in BioSciences* 8:155-165,1992; and Pearson *et al.*, *Meth. Mol. Bio.* 24: 307-331,1994. Altschul *et al.*, *Nat. Genet.* 6(2): 119-129, 1994 presents a detailed consideration of sequence alignment methods and homology calculations.

The NCBI Basic Local Alignment Search Tool (BLAST) (see Altschul *et al. J. Mol. Biol.* 215: 403-410, 1990) is available from several sources, including the National Center for Biotechnology Information (NCBI, Bethesda, MD) and on the Internet, for use in connection with the sequence analysis programs blastp, blastn, blastx, tblastn and tblastx. The Search Tool can be accessed at the NCBI website, together with a description of how to determine sequence identity using this program.

Nucleic acid sequences that do not show a high degree of identity can nevertheless encode similar amino acid sequences, due to the degeneracy of the genetic code. It is understood that changes in nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid molecules that all encode substantially the same protein.

Serial analysis of gene expression (SAGE) is the use of short diagnostic sequence tags to allow the quantitative and simultaneous analysis of a large number of transcripts in tissue, as described in Velculescu *et al.*, *Science* 270:484-487, 1995.

A **standard** is a reference against which a value (*e.g.*, level of expression of a marker) can be compared. By way of example, a non-cancerous cell line may be used as a standard for comparing the level of expression of tumor markers in an ovarian tumor sample. Non-limiting examples of standards useful with the disclosed methods of analysis of patterns of expression of markers include a non-cancerous sample (*e.g.*, normal ovarian tissue), a sample from a subject prior to development of a cancer or at an earlier stage of the cancer, and a cell line (*e.g.*, immortalized ovarian epithelial cells, such as IOSE cells) considered to display wild-type expression levels of the markers. In some embodiments, a reference RNA is arbitrarily chosen, but used consistently in relation to all tumor samples.

A **subject** is a living multi-cellular vertebrate organisms, a category that includes both human and non-human mammals.

A **therapeutic agent,** as used in a generic sense, is a composition used for treating a subject, such as a pharmaceutical or prophylactic agent.

A **transformed** cell is a cell into which has been introduced a nucleic acid molecule by molecular biology techniques. As used herein, the term transformation encompasses all techniques by which a nucleic acid molecule might be introduced into such a cell, including transfection with viral vectors, transformation with plasmid vectors, and introduction of naked DNA by electroporation, lipofection, and particle gun acceleration.

Treating a disease includes inhibiting or preventing the partial or full development or progression of a disease (*e.g.*, ovarian cancer), for example in a person who is known to have a predisposition to a disease. An example of a person with a known predisposition is someone having a history of breast or ovarian cancer in his or her family, or who has been exposed to factors that predispose the subject to a condition, such as exposure to radiation. Furthermore, treating a disease refers to a therapeutic intervention that ameliorates at least one sign or symptom of a disease or pathological condition, or interferes with a pathophysiological process, after the disease or pathological condition has begun to develop.

In some aspects, a more aggressive treatment may be selected if warranted. By way of example, if a subject is found to have a *BRCA1*-like or *BRCA2*-like gene expression pattern, a more aggressive treatment, such as chemotherapy, radiotherapy, or surgical removal of the affected tissue and/or surrounding area may be selected.

A **tumor** is an abnormal mass of tissue, or neoplasm that may be either malignant or non-malignant. "Tumors of the same tissue type" refers to primary tumors originating in a particular organ (such as breast, ovary, bladder or lung). Tumors of the same tissue type may be divided into tumor of different sub-types, for example ovarian carcinomas can be further classified based on tumor histology as adenocarcinoma, serous, endometrial, clear cell or mixed. Tumors may also be classified according to a genetic abnormality associated with the development of that type of tumor. By way of example, a tumor associated with a defect in tumor suppressor genes *BRCA1* or *BRCA2* is referred to herein as a "*BRCA1*- or *BRCA2*-linked" tumor. As described herein, a **sporadic ovarian tumor** is a tumor arising for a reason other than a mutation in *BRCA1* or *BRCA2*. However, the similarities in the pattern of expression of ovarian cancer markers in sporadic tumors to those in *BRCA1*-linked and *BRCA2*-linked tumors can be used to classify sporadic tumors into "*BRCA1*-like" or "*BRCA2*-like" tumors, using the methods of the disclosure. A "non-BRCA-type" tumor is one that has a pattern of expression of ovarian cancer markers unlike a BRCA1-like or BRCA2-like tumor.

A **vector** is a nucleic acid molecule as introduced into a host cell, thereby producing a transformed host cell. A vector may include nucleic acid sequences that permit it to replicate in the host cell, such as an origin of replication, and may also include one or more therapeutic genes and/or selectable marker genes and other genetic elements known in the art. A vector can transduce, transform, or infect a cell, thereby causing the cell to express nucleic acids and/or proteins other than those native to the cell. A vector optionally includes materials to aid in achieving entry of the nucleic acid into the cell, such as a viral particle, liposome, protein coating or the like.

Unless otherwise explained, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The singular terms "a," "an," and "the" include plurals unless the context clearly indicates otherwise. Similarly, the word "or" is intended to include "and" unless the context clearly indicates otherwise. "Comprises" means "includes." It is further to be understood that all base sizes or amino acid sizes, and all molecular weight or molecular mass values, given for nucleic acids or polypeptides are approximate, and are provided for description. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including explanations of terms, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

III. Description of Several Specific Embodiments

Provided herein are methods of diagnosing or prognosing development or progression of ovarian cancer in a subject, which methods involve detecting altered expression of at least one marker (*e.g.*, a nucleic acid molecule such as one listed in Table 1 or genes, cDNAs or other polynucleotide molecules comprising one of the listed sequences, or a fragment thereof, or a protein, such as one encoded by such a nucleic acid molecule, or fragment of such protein). In certain embodiments, altered expression is detected in more than marker, for instance in at least 50, at least 100, at least 200, or at least 400 or more nucleic acid molecules listed in Table 1, or encoded for by a nucleic acid molecule listed in Table 1. In certain specific embodiments, no more than the molecules listed in Table 6, Table 7, Table 8, Table 9, Table 10 or Table 11 are included in such analysis.

Additionally provided herein are methods for the classification of ovarian tumors as *BRCA1*-like, *BRCA2*-like or non-BRCA-like tumors based upon expression profiles of selected markers. Using the expression profile data, multiple types of comparisons can be made to provide qualitative and quantitative information about the tumor type. Non-limiting examples of such comparisons include visual examination of color profiles of hierarchically clustered markers on a cDNA microarray, multidimensional scaling to the determine relative distance of the analyzed markers, and compound covariate prediction analysis to statistically classify a given tumor into one of two classes based upon the logarithmic expression ratio of the expression of at least one known classifying marker. In a specific non-limiting example, logarithmic expression ratios are generated and used to classify tumor types by comparing to markers known to have a logarithmic expression ratio associated with *BRCA1*-like, *BRCA2*-like or non-BRCA-like tumors (see Example 4).

Also encompassed herein are arrays containing two or more disclosed markers. Certain of such arrays are nucleic acid arrays that contain at least one marker, for instance at least one or more, such as 5, 10, 15, 25, 50, 100, 150, 200, 250, 300, 350, 400 or more nucleic acid molecules listed in Table 1 (or genes, cDNAs or other polynucleotide molecules comprising one of the listed sequences, or a fragment thereof), or a fragment of such protein, or an antibody specific to such a protein or protein fragment. Such arrays can also contain any particular subset of the nucleic acids (or corresponding molecules) listed in Tables 1-11 or all of those nucleic acids. Certain arrays (as well as the methods described herein) also may include nucleic acid molecules that are not listed in Table 1.

Certain of the encompassed methods involve measuring an amount of the ovarian cancer-related molecule in a sample (such as a serum or tissue sample) derived or taken from the subject, in which a difference (for instance, an increase or a decrease) in level of the ovarian cancer-related molecule relative to a standard such as a sample derived or taken from the subject at an earlier time, is diagnostic or prognostic for development or progression of ovarian cancer.

In some embodiments, altered expression of ovarian cancer-related nucleic acid molecules is detected using, for instance, *in vitro* nucleic acid amplification and/or nucleic acid hybridization. The

results of such detection methods can be quantified, for instance by determining the amount of hybridization or the amount of amplification of the nucleic acid molecules.

Specific embodiments of methods for detecting altered expression of at least one ovarian cancer-related molecule use the arrays disclosed herein. Such arrays may be nucleotide (*e.g.*, polynucleotide or cDNA) or protein (*e.g.*, peptide, polypeptide, or antibody) arrays. In such methods, an array may be contacted with polynucleotides or polypeptides (respectively) from (or derived from) a sample from a subject. The amount and/or position of expression of the subject's polynucleotides or polypeptides then can be determined, for instance to produce a gene expression profile for that subject. Such gene expression profile can be compared to another gene expression profile, for instance a control gene expression profile from a subject having a known ovarian cancer-related condition. Similarly, protein arrays can give rise to protein expression profiles. Both protein and gene expression profiles can more generally be referred to as expression profiles. Expression profile data can be used to generate logarithmic expression ratios for use in compound covariate prediction analysis.

Other embodiments are methods that involve providing nucleic acids from the subject; semi-quantitatively amplifying the nucleic acids to form nucleic acid amplification products using primers; quantifying the amount of the nucleic acid amplification products; and comparing results to expression levels obtained using cDNA microanalysis. The sequence of such primers may be selected to bind specifically to a nucleic acid molecule listed in Table 1, or a nucleic acid molecule represented by those listed in Table 1. In specific examples of such methods, the primers are selected to amplify a nucleic acid product encoding topoisomerase II (*TOP2A*) (SEQ ID NO: 448), regulator of G-protein signaling 1 (RGS1) (SEQ ID NO: 398), invariant gamma-chain-associated protein (CD74) (SEQ ID NO: 89-91), epididymis-specific, whey-acidic protein (HE4) (SEQ ID NO: 60), major histocompatibility complex, class II, DR beta 1 protein (HLA-DRB1) (SEQ ID NO: 87-88), or zinc finger protein (ZFP36) (SEQ ID NO: 167-168).

Also encompassed are methods of ovarian cancer therapy, in which classification of a tumor of a patient into a *BRCA1*-like, *BRCA2*-like or non-BRCA-like tumor type aids in the selection of a treatment regimen. In some examples, the treatment selected is specific and tailored for the subject, based on the analysis of that subject's profile for one or more ovarian cancer-related molecules.

Other embodiments are kits for classifying tumors into a *BRCA1*-like, *BRCA2*-like or non-BRCA-like tumor class, which kits may include a binding molecule that selectively binds to the marker that is the target of the kit. In some examples of such kits where the marker is an ovarian cancer-related protein, the binding molecule provided in the kit may be an antibody or antibody fragment that selectively binds to the target ovarian marker protein. In other examples of such kits where the ovarian cancer-related marker level is a nucleic acid, the binding molecule provided in the kit may be an oligonucleotide capable of hybridizing to the nucleic acid marker molecule.

Further embodiments are methods of screening for a compound useful in treating, reducing, or preventing ovarian cancer or development or progression of ovarian cancer. Such methods involve determining if a test compound alters the gene expression profile of a subject (or cells of an *in vitro*

assay) so that the profile more closely resembles a wild-type expression profile than it did prior to such treatment, and selecting a compound that so alters the gene expression profile. In specific examples of such methods, the test compound is applied to a test cell. In some of such methods, the profile is determined or measured in an array format.

Also encompassed are compounds selected using the methods described herein, which are useful in treating, reducing, or preventing ovarian cancer or development or progression of ovarian cancer.

Also disclosed herein are uses of identified target ovarian cancer-related molecules for the development of antibodies, including therapeutic antibodies that affect an ovarian cancer-related pathway. It is also envisioned that the disclosed ovarian cancer-related molecules can be used as vaccines, for instance as "cancer vaccines" to elicit an immune response from a subject that renders the subject more resistant to developing or progressing through a stage of ovarian cancer.

IV. Gene expression profiling of ovarian cancer tumor tissue using disclosed markers

The present disclosure concerns gene expression profiling of ovarian tumor tissue from a subject for use in diagnosing, prognosing, staging, preventing, and treating the disease. Measurement of expression of genes within a tissue sample provides information regarding proteins that may be active during cancer mechanisms. Hence, the gene expression profile of tumor tissue may be compared against the profile for known markers for ovarian cancer, such as those disclosed herein (see Table 1).

Using the gene expression profile, an ovarian tumor from a subject may be classified into a *BRCA1*-like, *BRCA2*-like, or non-*BRCA*-like tumor. Because the prognosis for a patient having a *BRCA1* or *BRCA2* mutation is poorer than for patient having non-BRCA-like mutations, classification of tumors into these groups is helpful in selecting treatment strategies and aids a clinician in deciding whether to employ a more aggressive regimen in treating the patient, for instance radiotherapy, chemotherapy, or surgical removal of the affected tissue. In addition, classification of a sporadic tumor into a *BRCA1*-like or *BRCA2*-like classification may provide similar guidance in treating the patient. For example, a subject who has a *BRCA1*-like or *BRCA2*-like sporadic tumor may be treated similarly to a subject who has a *BRCA1*-linked or *BRCA2*-linked tumor. The identification of *BRCA1*- or *BRCA2*-like sporadic tumors also allows tumors (or subjects) to be selected for specific drug regimens that are particularly effective with the associated mutation type.

The ovarian cancer-linked markers disclosed herein are believed to be useful as diagnostic or prognostic indicators of *BRCA1*-like, *BRCA2*-like and non-*BRCA*-like ovarian cancer. In addition, the markers are believed to be useful in applications for treating ovarian cancer as the basis of new therapeutic targets, for the development of new anti-cancer therapeutic compounds, and/or to select particularly appropriate existing treatments. For example, the expression levels of these markers can be examined to monitor the effectiveness of anti-cancer treatments where an increase in or decreased level of nucleic acid expression opposite of the ovarian cancer-indicative pattern disclosed herein

indicates an effective anti-cancer treatment. Beyond use in generating an expression profile, certain of the identified genes or EST sequences provided herein are believed to have individual use as cancer markers.

A. ***Generating gene expression information and logarithmic expression ratios.***

cDNA microanalysis allows for simultaneous analysis of the expression of multiple genes within various tissue samples, and is therefore useful in generating gene expression profiles. To perform cDNA microarray analysis, RNA is isolated from a subject and cDNA is synthesized from the RNA according to standard methods (see Sambrook *et al.*, *Molecular cloning, a laboratory manual*. 2nd ed. Cold Spring Harbor Laboratory, Cold spring Harbor, New York, 1989). Relative over-expression of the mRNA in the cancerous tissues can be measured against non-cancerous reference baselines (*e.g.*, ovarian tissue from a subject not having ovarian cancer or an ovarian cell line, such as an immortalized ovarian cell line), to provide a framework for determining normal expression versus altered expression (genes that are either overexpressed or underexpressed). Nucleic acids that are overexpressed may be used as markers for ovarian cancer, while genes that are underexpressed may be putative tumor suppressors.

cDNA microarrays containing 7,651 sequence-verified features were constructed and applied to analyze the mRNA expression profile of sixty-one subjects with pathologically-confirmed epithelial ovarian adenocarcinoma having matched clinicopathologic features (see Alizadeh *et al.*, *Nature* 403: 503-511, 2000; Perou *et al.*, *Nature* 406: 747-752, 2000; Bubendorf *et al.*, *J Natl. Cancer Inst.*, 91(20): 1758-64, 1999; Welsh *et al.*, *Proc Natl Acad Sci. U S A* 98: 1176-1181, 2001). These included eighteen cases linked to *BRCA1* founder mutations, sixteen cases linked to *BRCA2* founder mutations, and 27 cases negative for any founder mutations (*e.g.*, sporadic ovarian epithelial cancer). These samples were compared to expression levels of these same features in an immortalized normal ovarian surface epithelium cell line (IOSE). Statistical tools, including a modified F-test with $P<0.0001$ considered significant (*e.g.*, a 99.99% confidence level) were then used to analyze the data (*e.g.*, to differentiate gene expression profiles associated with ovarian cancer), enabling a comprehensive, genomics-based analysis of the mRNA expression profiles of these ovarian cancer subjects.

The logarithmic expression ratios for the spots on each array were normalized by subtracting the median log ratio for the same array. Data were filtered to exclude spots with size less than 25 μm, intensity less than two times background or less than 300 units in both red and green channels, and any flagged or missing spots. In addition, any features found to be missing or flagged in greater than 10% of the arrays were not included in the analysis. Application of these filters resulted in the inclusion of 6,445 of the total 7,651 features in subsequent analyses. Statistical comparison between tumors groups was performed using the "BRB Array Tools" software (developed by Dr. Richard Simon and Amy Peng, Biometrics Research Branch, Division of Cancer Treatment and Diagnosis, NCI, USA), consisting of a modified F-test with $P < 0.0001$ (99.99% confidence level) considered significant (see Example 4). This stringent *P* value is selected in lieu of the Bonferroni correction for

multiple comparisons, which was deemed excessively restrictive (*see* Bland and Altman, *B.M.J.*; 310: 170, 1995). See Example 4 for an example of how an ovarian tumor is analyzed using the disclosed methods.

In addition to statistical analysis, multidimensional scaling (MDS) and hierarchical clustering techniques using a correlation metric and average linkage were used for evaluating overall gene expression. Using these techniques, a large set of genes and other encoding sequences (*e.g.*, expressed sequence tags, ESTs) have been identified (Table 1), the expression of which varies in subjects having ovarian cancer (see Addendum). Other confidence levels could be used to select ovarian cancer-related molecules, such as 98%, 95%, 90%, 85%, and so forth (see Jain *et al.*, *IEEE Transacations on Pattern Analysis and Machine Intelligence* 22(1): 4-37, 2000). Molecules identified as being linked to ovarian cancer (referred to generally herein as ovarian cancer-related molecules) using the methods described herein can be arranged on arrays for use in diagnostic and prognostic methods. Specific arrays are contemplated that are constructed using molecules identified at differing confidence levels. Specific examples of such arrays include arrays that detect altered expression of at least 2, 5, 10, 20, 30, or 50 of these molecules.

B. ***Comparison of ovarian epithelial adenocarcinoma cells to immortalized ovarian surface epithelium cells.***

In a comparison of ovarian epithelial adenocarcinoma cells to immortalized ovarian surface epithelium cells, the largest contrast in gene expression was observed between *BRCA1*- and *BRCA2*-linked tumors, with multiple genes showing significant differences in expression levels. This group of genes was also able to segregate the sporadic tumors into two major "*BRCA1*-like" and "*BRCA2*-like" subgroups, indicating that *BRCA*-related pathways are also involved in sporadic ovarian cancers. In addition, two previously unreported gene expression patterns were noted. First, six of the genes differentially expressed between *BRCA1*-linked and sporadic tumors map to Xp11.23 and all exhibited higher mean expression levels in the *BRCA1*-linked samples [*WAS* (SEQ ID NO: 524-526), *EBP* (SEQ ID NO: 529), *SMC1L1* (SEQ ID NO: 529), *PCTK1* (SEQ ID NO: 527-528), *ARAF1* (SEQ ID NO: 531-532), and *UBE1* (SEQ ID NO: 533), see Figure 3]. Second, compared to immortalized ovarian surface epithelium cells, several interferon-inducible genes were noted to be overexpressed in the majority of all tumor samples [*SIAT1* (SEQ ID NO: 73), *TNFSF10* (SEQ ID NO: 104-106), *ABCB1* (SEQ ID NO: 164-166), *CP* (SEQ ID NO: 83-84), *HLA-DRB5* (SEQ ID NO: 85-86), *HLA-DRB1* (SEQ ID NOS: 87-88, 100, 101-103), *CD74* (SEQ ID NO: 92-93), *HLA-DRA* (SEQ ID NO: 94-96), *HLA-DPA* (SEQ ID NO: 97-99), *IFITM1* (SEQ ID NOS: 50-51, 52-54), *IFITM2* (55-57, 58-59), *A2M* (SEQ ID NO: 193-195), *G1P3* (68-69), *IGKC* (SEQ ID NOS: 112-114, 115-116), *SCYB10* (SEQ ID NO: 120-121), *Col3A1* (SEQ ID NO: 141-143), *HLA*-B (SEQ ID NO: 154-156), and *HLA-C* (SEQ ID NO: 157-159), see Figure 4]. In terms of overall differential gene expression, *BRCA1* and *BRCA2*-linked tumors express genes more different from each other than from sporadic (non-*BRCA*-linked) tumor samples.

The identified ovarian cancer-related genes represent putative mediators of ovarian cancer, and as such are candidate targets for the development of novel therapeutics for the treatment of ovarian cancer using conventional techniques. By way of example, a candidate drug, targeted at restoring expression of a gene of the disclosure, could be examined using cDNA microarray analysis for utility in influencing growth of ovarian cancer cells. Thus, use of cDNA microarray techniques for genomics-based discovery of genes variably expressed during ovarian cancer provides for the identification of novel therapeutic targets for treatment of ovarian cancer.

It is contemplated that certain of the ovarian cancer markers identified herein encode or correspond to soluble proteins, while others encode or correspond to membrane associated or membrane integral proteins, some of which are exposed at least to a certain extent on the exterior of a cell in which they are expressed. In some embodiments, those ovarian cancer-related molecules that are expressed at or on the surface of a cell are selected as therapeutic targets, for instance for targeting with an antibody-based therapy, which is facilitated by the access of the ovarian cancer-related molecule to the extracellular matrix. These ovarian cancer markers may be described as being "drug accessible." In addition, such soluble ovarian cancer markers, if secreted, may be detected in a blood or serum sample from the subject.

C. ***Comparison of ovarian epithelial cancer cells to normal postmenopausal ovarian samples.***

cDNA microarrays containing 7,600 sequence-verified features were constructed and applied to analyze the mRNA expression profile of 61 subjects with ovarian epithelial cancer as compared to two normal postmenopausal ovarian samples.

Gene expression in each sample (normal or tumor) was directly compared to a "reference RNA" consisting of a mix of nine different human cell lines (breast adenocarcinoma, hepatoblastoma, cervical adenocarcinoma, testicular embryonal carcinoma, glioblastoma, melanoma, liposarcoma, histiocytic lymphoma, T cell lymphoblastic leukemia, and plasmacytoma/myeloma, Stratagene, La Jolla, CA). The raw gene expression data was used to calculate the logarithmic expression ratio for each gene. The logarithmic expression ratios ("log ratio") obtained from this comparison were then normalized and statistically compared to one another, providing for indirect comparison of gene expression in tumors and normal ovarian samples. This was accomplished by scoring the magnitude of differential expression of each gene (between normal and cancer samples) according to the formula:

$$\frac{\text{(average cancer log ratio – average normal log ratio)}}{\text{(standard deviation cancer + standard deviation normal)}} = \text{magnitude of differential expression}$$

In genes showing a large mean expression difference between normal and cancerous samples, the magnitude of differential expression has a greater value, while the intra-group variability in expression ratios is low.

Genes were then ranked according to the magnitude of differential expression and the highest-ranking genes were considered to be the best candidates for differentiating normal from malignant ovarian samples (see Furey *et al.*, *Bioinformatics* 16(10): 906-14, 2000).

Using these techniques, a large set of genes and other encoding sequences (*e.g.*, ESTs) that are under-expressed in subjects having ovarian cancer have been identified (see Table 4). These under-expressed ovarian cancer markers represent putative tumor suppressors, and as such are candidate targets for the development of novel therapeutics for the treatment of ovarian cancer using conventional techniques. By way of example, induction of expression of one or more of these markers through therapeutic means (*e.g.*, induction by a drug or gene therapy) may inhibit tumor growth and/or increase tumor cell death, for instance through stimulation of apoptotic pathways.

Furthermore, a large set of genes and other encoding sequences (*e.g.*, ESTs) have been identified (see Table 5), the expression of which is overexpressed in subjects having ovarian cancer. These overexpressed ovarian cancer markers represent putative mediators of ovarian cancer, and as such are candidate targets for the development of novel therapeutics for the treatment of ovarian cancer using conventional techniques. Over-expression of one or more such markers can also be detected in the body (for example using a serum test to detect or monitor progression of ovarian cancer).

In addition, six of the markers identified herein (*e.g.*, *WAS* (SEQ ID NO: 524-526), *PCTK1* (SEQ ID NO: 527-528), *UBE1* (SEQ ID NO: 533), *SMCIL1* (SEQ ID NO: 529), *ARAF1* (SEQ ID NO: 531-532), and *EBP* (SEQ ID NO: 529)) have all been mapped to chromosome Xp11 (see Example 1). Hence, this chromosome could contain additional genes and ESTs that may be useful as markers for prognosing, diagnosing and monitoring ovarian cancer. The methods of the disclosure can be used to find additional genes and ESTs in this region for use as ovarian cancer markers.

V. Methods of Classifying Tumors into Subgroups

Disclosed herein are multiple methods of classifying tumors into subtypes based upon the expression of disclosed ovarian tumor markers (see Table 1).

A. Comparison of raw expression data

The expression data of one or more ovarian cancer markers can be compared between samples and analyzed to detect differences in expression between the markers. The expression of an individual marker can be stated in ratio or "fold" form relative to the expression of the standard. For instance, in Table 4, the average logarithmic ratio of the gene expression for the standard ("normal") for *ITM2A* (SEQ ID NO: 202) is 1.145, while the average logarithmic ratio of the gene expression in cancer cells was -2.036. These numbers can be compared to derive a value for the difference in expression by calculating the expression ratio of each number, and dividing the expression ratio for the average log cancer value by the expression ratio for the average log normal value. Hence:

Expression ratio of average log normal: $2^{1.145} = 2.211$

Expression ratio of average log cancer $2^{-2.036} = 0.244$

Ratio (cancer to normal) = (0.244)/(2.211) = 0.110

Thus, *ITM2A* is under-expressed in cancer by a ratio of 0.110 to 1 (*i.e.,* in ovarian cancer tissue, *ITM2A* expresses at approximately 10% of the expression level seen in wild-type cells).

Collections of such data can be assembled to provide a gene expression profile, as discussed above. With such profiles, the standard deviation of the expression ratio of each gene can be measured by obtaining the square root of the variance of the expression data as described by Jaccard and Becker (in *Statistics for the Behavioral Sciences*, 2nd ed., Wadsworth Publishing Co., Belmont, California, 1990) and Myers and Well. (in *Research Design and Statistical Analysis*, University of Massachusetts, Amherst, Massachusetts, 1995).

Further analysis can include a Student's t-test, to determine if the mean expression of two groups (*e.g.*, *BRCA1*-like and non-*BRCA*-like, *BRCA2*-like and non-*BRCA*-like, etc.) are statistically different from each other.

Due to the range over which genes may express, it may be useful to perform statistical analyses using the logarithmic expression value for each marker (see Example 1). However, calculations using the logarithmic expression values may dilute the ability of certain analyses to determine differences. Hence, it may be useful to employ multiple methods of analysis to ascertain relative values in expression (see Jain *et al.*, *IEEE Transacations on Pattern Analysis and Machine Intelligence* 22(1): 4-37, 2000).

B. Visual analysis of hierarchical clustering

Methods disclosed herein include hierarchical clustering analysis of genes with statistically significant differential expression between sets of tumor groups. Hierarchical clustering can be used to cluster objects (*e.g.*, genes, such as the ovarian cancer markers listed in Table 1) to represent relationships among the objects. The relationships are represented, for example by a tree whose branch lengths reflect the degree of similarity between the objects (see *e.g.*, Figure 2B).

Optionally, hierarchical clustering can be combined with a graphical representation of the primary data by representing each data point with a color that quantitatively and qualitatively reflects the original experimental observations. The use of color representations, along with statistical organization, provides a graphical display that provides visual information about expression of the genes. Hence, the methods disclosed herein can provide visual information regarding degrees of similarity (*e.g.*, patterns of under-expression or over-expression) between assessed genes in different samples, for instance in samples of *BRCA1*-linked, *BRCA2*-linked and sporadic ovarian tumor samples (see Figure 2B).

At the first iteration, each object is considered to be its own group, and the pair of objects with the smallest distance between them is merged into a new group. Each subsequent iteration

merges two groups to form a new group, until finally all objects end up merged into a single group. The classification tree, or dendrogram, graphically represents the sequence of clusters formed at each iteration of merges, as well as the distance between clusters at each merge (here, Figure 2). This technique is widely employed to represent gene expression information obtained from microarray experiments (see Eisen *et al., Proc. Natl. Acad. Sci. U.S.A.* 95(25): 14863-8, 1998).

The gene expression data disclosed herein were analyzed by calculating the Pearson correlation coefficient to obtain a gene expression similarity metric. To describe, *Gi* is (log-transformed) primary gene expression data for gene *G* in each tumor sample, represented as variable *i*. For any two genes *X* and *Y* observed over a series of *N* tumor samples, a similarity score can be computed as follows:

$$S(X, Y) = \frac{1}{N} \sum_{i=1,N} \left(\frac{X_i - X_{offset}}{\Phi_X}\right)\left(\frac{Y_i - Y_{offset}}{\Phi_Y}\right)$$

where

$$\Phi_G = \sqrt{\sum_{i=1,N} \frac{(G_i - G_{offset})^2}{N}}.$$

When G_{offset} is set to the mean of the gene expression levels of the tumor samples for gene *G*, then Φ_G becomes the standard deviation of *G*, and *S(X, Y)* is exactly equal to the Pearson correlation coefficient of the gene expression levels for genes *X* and *Y*. Values of G_{offset} that are not the average of the gene expression levels for gene *G* are used when there is an assumed unchanged or reference state (*e.g.*, the gene is not over-expressed or under-expressed) represented by the value of G_{offset}, against which changes are to be analyzed; in all of the examples presented here, G_{offset} is set to 0, corresponding to a fluorescence ratio of 1.0.

By way of example, Figures 2A and 2A' demonstrate that expression of the disclosed markers can be used to visualize different tumor types. Hierarchical clustering was performed using with a Pearson correlation metric and average linkage were used for evaluating overall gene expression for the sixty-one *BRCA1*-linked, *BRCA2*-linked and sporadic tumors (see Example 1). When applicable, all statistical tests were two-sided.

In Figure 2, B2 represents *BRCA2*-linked tumors, and B1 represents *BRCA1*-linked tumors. The red and green intensities represent standard normal deviation (Z score) values from each marker's means expression level (represented as black) across the sixty-one tumors samples. Red represents increased expression and green represents decreased expression. The differences in gene expression can be appreciated by looking at the groupings apparent in Figure 2A. The genes in the left half of the Figure 2A are from *BRCA2*-linked tumors and the genes in the right half are from *BRCA1*-linked tumors. As can be seen with casual observation, gene expression between these two tumor groups differs relative to the control (IOSE cells). Specifically, *BRCA2*-linked tumors contain under-expressing genes that correlate to these genes in the upper left and lower right quadrants of

Figure 2A, which are represented as primarily green in color. Furthermore, the genes in the upper right and lower left quadrant, which are represented as primarily red in color, correlate to genes that are generally over-expressed relative to the control IOSE cells. Hence, hierarchical clustering can be used to qualitatively visualize differences in the expression patterns of samples.

C. Multidimensional Scaling

Multidimensional scaling is a dimension reduction procedure that can be used for visualization purposes. Each experiment can be represented by its expression profile, which is a K-dimensional vector of log-ratios, where K is the number of clones represented after filtering. The multidimensional scaling procedure reduces each experiment's expression profile from K-dimensional space to 3-dimensional space, by attempting to preserve distances between the N experiment vectors. The distance metric needs to be specified when using the multidimensional scaling tool. First, the N x N distance matrix is computed, which quantifies the relationships between the N experiments in the series of chips. For each of the N vectors in K-dimensional space, the multidimensional scaling procedure finds a vector in 3-dimensional space, such that the N x N distance matrix computed in 3-dimensional space approximates the N x N distance matrix computed in K-dimensional space. The relationships between the N experiments can then be visualized by plotting the N vectors in 3-dimensional space, in which each of the N points represents a single experiment. A rotating 3-dimensional visualization tool can be used for discovery of experiment clusters.

By way of example, the gene expression data of 6445 filtered genetic elements of the sixty-one ovarian tumor samples (see Example 1) was used in multidimensional scaling to generate a 3-D diagram for visualization of the respective differences between the expression patterns of each tumor sample. As seen in Figure 1, the data segregate into different areas of the 3-D space based on similarities in gene expression within the tumor type. In particular, the *BRCA1*-linked tumors (dark circles) segregate higher into the cube than the *BRCA2*-linked tumors (open circles). The sporadic tumor samples (asterisks) also fell into higher and lower areas of the cube, indicating that they segregate into *BRCA1*-type and *BRCA2*-type expression patterns. Thus, multidimensional scaling can be used to make a qualitative distinction regarding the expression patterns of these samples.

Multidimensional scaling can be used to qualitatively assess the expression pattern of an unknown tumor type. Expression data for a plurality of *BRCA1*-type and *BRCA2*-type markers is generated using the tumor tissue (for instance, on a cDNA microarray) relative to a standard ovarian tissue (*e.g.*, from a subject not having ovarian cancer, immortalized ovarian epithelial cells, etc.), and logarithmic ratios of the gene expression data are calculated. To compare the pattern of expression of the plurality of the known *BRCA1*-type and *BRCA2*-type markers to the unknown ovarian sample, the K-dimensional vectors of the logarithmic expression ratios for all expression data are calculated as discussed above. Next, the K-dimensional vectors are plotted in a 3-dimensional space and the layout of the data compared. Similar to Figure 1, the unknown sample data should cluster either near the *BRCA1*-like or *BRCA2*-like tumors, or alone (which would indicate that it is a non-BRCA-like

tumor). Hence, multidimensional scaling can be used to make a qualitative distinction regarding the expression patterns of an unknown samples in comparison to known *BRCA1*-type and *BRCA2*-type markers. In addition, more than one unknown sample can be used in this analysis.

D. Compound Covariate Predictor Analysis

Segregation into tumor types can be performed using compound covariate predictor analysis, which creates a multivariate predictor for one of two classes to each sample (see Example 4). Markers included in the multivariate predictor are those that are univariately significant at the selected significance cutoff (*e.g.*, P<0.0005). The multivariate predictor is a weighted linear combination of log-ratios (or log intensities for single-channel arrays) for genes that are univariately significant. The weight consists of the univariate t-statistics for comparing the classes, and is calculated using the equation:

$$CCP = \Phi_i t_i *(x_i - m_i)$$

where t_i = t-value for gene *i* (see Table 10), x_i = logarithmic ratio of the gene expression (*i*) in the new sample to be classified, and m_i = midpoint between the two classes for gene *i* (see Table 10). The index *i* runs over all the genes that are significant in the original analysis (*i.e.* all 62 genes in Table 10). If the log ratio x_i is missing for gene *i* in the new sample to be classified, then it should be assigned as m_i for that gene, to cause the result of the calculation to be zero for that gene. If the compound covariate predictor value is positive, then the tumor classified as one of the first type (*e.g.*, BRCA1-like). If the compound covariate predictor value is negative, then the tumor is classified as belonging to the second type (*e.g.*, BRCA2-like).

A second method of tumor classification using a compound covariate predictor model can be found in Radmacher *et al.*, "A paradigm for class prediction using gene expression profiles," found on the National Cancer Institute Internet website. This publication is expressly incorporated by reference herein.

In order to determine whether a tumor is classified as *BRCA1*-like or *BRCA2*-like using a single markers in Table 10, the following steps are used:

1. Gene expression information is obtained, for instance on a cDNA microarray, using the same standard (*e.g.*, IOSE cells) that was used to obtain the marker data.

2. The gene expression data is converted into a logarithmic ratio using log base 10. Hence, a tumor that has a gene expression value for gene *KIAA00008* of 0.45 would have a log base 10 ratio of -0.346.

3. The midpoint value of Table 10 is subtracted from the logarithmic ratio, and multiplied by the t-value for *KIAA00008* for the data set. Thus,

[(-0.346) – (-.431)] * (-8.0421) = -0.51.

The values for the average logarithmic ratio for *BRCA1*-linked and *BRCA2*-linked values in the data set are then consulted. The obtained value will fall between the midpoint and one of these values because genes in which larger values of the logarithmic ratio are assigned to one class (*e.g.*, *BRCA1*-linked) will have weights of a value that is more negative with respect to the midpoint value (*e.g.*, -0.56864), whereas genes in which larger values of the logarithmic ratios are assigned to the other class (*e.g.*, *BRCA2*-linked) will have weights of a value that is more positive with respect to the midpoint value (*e.g.*, -0.29414). Hence, the obtained value, 0.1930, would fall on the more negative side of this data, and would therefore be classified as a *BRCA2*-like data set.

If this same analysis is performed using multiple markers, the method remains the same except that the data can be summed prior to performing the analysis. This method is a multivariate approach of the compound covariate analysis, and can be used to determine whether the pattern of expression of an unknown tumor is similar to a BRCA1-like or BRCA2-like pattern of expression.

Further analysis, such as a "leave-one-out" approach may additionally be employed to test the ability of the Compound Covariate Predictor to classify the tumors into additional subtypes, such as resistance to a therapeutic compound. See Radmacher *et al.*, "A paradigm for class prediction using gene expression profiles," found on the National Cancer Institute Internet website.

E. Comparisons using Databases

Due to the large amount of information associated with the analysis methods disclosed herein, it may be particularly useful to construct and/or consult databases of information for use in the analysis.

By way of example, the information generated by the methods of the disclosure can be stored in databases, such as a database of a plurality of markers known to express differently in *BRCA1*-like and *BRCA2*-like tumors (*e.g.*, Table 9). Such databases may be made publicly available, such as the Stanford Microarray Database (maintained by Stanford University, see Sherlock *et al.*, *Nucleic Acids Res.*, 29(1):152-155, 2001). These databases may be used to store reference data for use with the classification methods of the disclosure. In addition, such databases can be used to provide information regarding markers of potential use in diagnosing, prognosing, or monitoring ovarian cancer, for use by clinicians.

The use of databases to search for stored information is disclosed in U.S. 5,871,697 and 6,519,583 the methods of which are expressly incorporated herein.

VI. Kits for measuring the level or function of ovarian cancer-related molecules.

The nucleic acid sequences and ESTs disclosed herein can be supplied in the form of a kit for use in detection or monitoring ovarian cancer. In such a kit, one or more of the nucleic acid sequences and/or ESTs in Table 1 are provided in one or more containers, or in the form of a microarray. The kit may also contain reagents for use in preparing a biological sample of a subject for screening with the kit. The container(s) in which the reagent(s) and microarray(s) are supplied can be any conventional container that is capable of holding the supplied form, for instance, plastic

boxes, microfuge tubes, ampoules, or bottles. In some applications, negative controls obtained from a subject free from ovarian cancer may be provided in pre-measured (*e.g.*, single use) amounts in individual, typically disposable, tubes or equivalent containers. With such an arrangement, the sample to be tested for the presence of ovarian cancer can be added to the testing container and tested directly.

The amount/number of each testing reagent and container supplied in the kit can be any appropriate amount, depending for instance on the market to which the product is directed. For instance, if the kit is adapted for research or clinical use, the amount of each testing reagent and container provided would likely be an amount sufficient to screen several biological samples. Those of ordinary skill in the art know the amount of testing reagent that is appropriate for use in a single container. General guidelines may for instance be found in Innis *et al.* (*PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc., San Diego, CA, 1990), Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989), and Ausubel *et al.* (In *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1992).

A kit may include more than two nucleic acid sequences or ESTs, in order to facilitate screening of a larger number of ovarian cancer markers or tumor suppressors. For instance, the sequences set forth in Table 1, or a subset of (*e.g.*, 5, 10, 15, 20, 50, 100, 150, 200, 250, 300, 350, 400 or more) of these sequences, may be provided. By way of example, a provided subset could include the markers set forth in Table 6, Table 7, Table 8, Table 9, or Table 10. These sets of sequences are provided by way of example only, and are not intended to be limiting examples.

In some embodiments of the current disclosure, kits may also include the reagents necessary to carry out screening reactions, including, for instance, RNA sample preparation reagents, appropriate buffers (*e.g.*, polymerase buffer), salts (*e.g.*, magnesium chloride), and secondary detection reagents (*e.g.*, cyanine 5-conjugated dUTP).

Kits may in addition include either labeled or unlabeled sequences for use in detection of the expression levels.

Embodiments of the disclosure are illustrated by the following non-limiting Examples.

EXAMPLE 1

Identification of Genes with Altered Expression in Ovarian Cancer

This example describes how a first subset of the disclosed ovarian cancer-related nucleic acid molecules were identified. These ovarian cancer-related molecules show differences in expression in subjects having ovarian cancer compared to normal ovarian surface epithelial cells and are classified according to their *BRCA*-1, *BRCA*-2, and sporadic tumor status. The results of these studies have been published in Jazaeri *et al.*, *J. Natl. Cancer Inst.*, 94(13): 990-1000, 2002, which is incorporated by reference in its entirety herein.

Methods and Material:

Clinicopathologic characteristics of BRCA-*linked and sporadic ovarian cancers*: Sixty-one cases of pathologically-confirmed epithelial ovarian adenocarcinoma from the Memorial Sloan-Kettering Cancer Center were studied and screened for founder mutations. These included eighteen cases linked to *BRCA1*, sixteen cases linked to *BRCA2*, and twenty-seven sporadic cases. All patients were self-identified as Ashkenazi Jews and after informed consent underwent genotyping for germ-line founder mutations in *BRCA1* (185delAG and 5382insC) and *BRCA2* (6174delT) (see Boyd *et al.*, *JAMA*: 283: 2260-2265, 2000). Those cases with a *BRCA* mutation were categorized as having hereditary ovarian cancer and those without such a mutation as having sporadic ovarian cancer.

Tumor samples: In order to minimize confounding variables, *BRCA1*-linked, *BRCA2*-linked, and sporadic tumors of similar stage, grade, and histology were selected from the sixty-one individuals studied [18 *BRCA1* (185delAG, 5382insC), 16 *BRCA2* (6174delT), 27 sporadic tumors). The majority of tumors in all three groups were characterized by advanced stage, moderate to high grade (grade 2 or 3), and a predominance of serous histology. Hence, the clinicopathologic parameters of selected samples were well-matched and in agreement with those reported previously for these tumors types (see Boyd *et al.*, *JAMA*: 283: 2260-2265, 2000: Ramus *et al.*, *Genes Chrom. Cancer*: 25: 91-96, 1999).

All tumor samples had been flash frozen, embedded in OCT medium, and stored at -80° C. Isolation of RNA was performed using the RNeasy columns (Qiagen, Valencia, CA) according to the manufacturer's instructions. The integrity of RNA was verified by denaturing gel electrophoresis. Total RNA was linearly amplified using a modification of the Eberwine method (see Van Gelder *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 87: 1663-1667, 1990). Table 2 catalogs the clinicopathologic features of the tumor samples studied.

Table 2. Clinicopathologic features of tumor samples

Variable	*BRCA1*-linked	*BRCA2*-linked	Sporadic
Number of samples	18	16	27
Median Age* (SD)	50 (11)	60 (9)	69 (11)
Stage			
I	2 (11.1%)	0	0
II	0	2 (12.5%)	3 (11.1%)
III	11 (61.1%)	12 (75%)	24 (88.9%)
IV	5 (27.8 %)	2 (12.5%)	0
Grade			
1	0	0	0
2	4 (22.2%)	6 (37.5 %)	8 (29.6%)
3	14 (77.8%)	7 (43.8%)	16 (59.3%)
No.	0	3 (18.7%)	3 (11.1%)
Histology**			
Serous	9 (50%)	12 (75%)	16 (59.3%)
Endometrioid	3 (16.7%)	0	2 (7.4%)
Mucinous	0	0	0
Clear Cell	2 (11.1%)	0	0
Adenocarcinoma NOS	3 (16.7%)	3 (18.8%)	9 (33.3%)
other	1 (5.5%)	1 (6.2%)	0

* F test, P = .0002/ Data are the median +/- standard deviation.
** Chi test *P* value for differences in histology among tumor groups= 0.17
NOS = not otherwise specified

cDNA Microarrays: The cDNA microarrays consisted of 7,651 total features representing different (non-redundant) genes, and were manufactured at the National Cancer Institute microarray facility.

Total RNA was reverse-transcribed by using a 63 nucleotide synthetic primer containing the T7 RNA polymerase binding site (5'-GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG$(T)_{24}$-3' (SEQ ID NO: 1). Second strand cDNA synthesis (producing double-stranded cDNA) was performed with RNase H, *Escherichia coli* DNA polymerase I, and *E. coli* DNA ligase (Invitrogen, Carlsbad, CA). After cDNA was blunt-ended with T4 DNA polymerase (Invitrogen, Carlsbad, CA), it was purified by extraction with a mixture of phenol, chloroform, and isoamyl alcohol and by precipitation in the presence of ammonium acetate and ethanol. The double-stranded cDNA was then transcribed using T7 polymerase (T7 Megascript Kit, Ambion, Austin, TX), yielding amplified antisense RNA that was purified using RNeasy mini-columns (Qiagen, Valencia, CA). Pooled total RNA from two SV40 immortalized ovarian surface epithelial cell-lines (IOSE) was amplified and used as reference for cDNA microarray analysis.

Four μg of amplified RNA was reverse transcribed and directly labeled using cyanine 5-conjugated dUTP (tumor RNA) or cyanine 3-conjugated dUTP (IOSE RNA, provided by Dr. Jeff Boyd, Memorial Sloan-Kettering). Hybridization was performed in a solution of 5X SSC and 25%

formamide for 14-16 hours at 42°C. Slides were washed, dried, and scanned using an Axon 4000a laser scanner (Axon Instruments, Inc., Union City, CA).

Imaging and I.M.A.G.E. Analysis: Fluorescence intensities at the immobilized targets were measured by using an Axon GenePix Scanner and Genepix Pro 3.0 analysis software (Axon Instruments, Union City, CA). The raw data were then uploaded to a relational database maintained by the Center for Information Technology at the National Institutes of Heath. The cDNA clones are identified by their Integrated Molecular Analysis of Genomes and their Expression Consortium (I.M.A.G.E.) clone number.

Amplification of RNA: The first strand of RNA was synthesized, by adding 1-3 µg of total RNA into a reaction tube (*e.g.*, Eppendorf, or other container of suitable size), adding 1µl T7-$(dT)_{24}$ primer (2µg/µl), and bringing to a volume of 20 µl with nuclease-free water. The reaction was incubated at 70°C for 10 minutes, then spun briefly in a centrifuge and placed on ice. Four µl 5X first strand cDNA buffer was added, then 2 µl 0.1M DTT, 2µl 10mM dNTP mix (Amersham-Pharmacia, Piscataway, NJ 08855-1327 USA), 1µl RNasin (Promega, Madison, WI 53711 USA), and 2 µl Superscript II. The reaction was mixed well and incubated at 42°C for 1 hour. The tube was centrifuged briefly, and placed on ice. To synthesize the second strand, 91 µl DEPC-treated water was added, then 30 µl second strand buffer, 3 µl10 mM dNTP mix, 4 µl DNA Polymerase I (10U/µl), 1 µl DNA Ligase (10U/µl), and 1 µl RNAse H (2U/µl). The final reaction volume equaled 150 µl. Next, the tube was gently tapped to mix, then briefly centrifuged. The tube was incubated at 16°C for two hours, then 2 µl (10U) T4 DNA Polymerase was added. The tube was cooled for five minutes at 16°C, then the reaction stopped with 10 µl of 0.5 M EDTA. Ten µl of 1M NaOH were added, then the reaction was incubated at 65°C for 10 minutes. The solution was neutralized by addition of 25 µl Tris-HCl (pH=7.5).

Clean Up of Double Stranded cDNA: The Phase Lock Gel (PLG) was pelleted in a microcentrifuge at maximum speed for 30 seconds. 198 µl (equal volume) of (25:24:1) Phenol : chloroform : isoamyl alcohol (saturated with 10 mM Tris-HCl pH 8.0/1 mM EDTA) was added to the final DNA synthesis preparation (198µl) to a final volume of 396 µl. The solution was mixed well by pipetting up & down vigorously. The entire cDNA-phenol/chloroform mixture was transferred to the PLG tube, and microcentrifuged at maximum speed for two minutes. The aqueous supernatant was transferred to a new 1.5 ml tube, and 1 µl linear acrylamide was added. 0.5 volumes of 7.5M Ammonium Acetate + 2.5 volumes (include the added Ammonium Acetate) of 95% ethanol stored at –20 to the sample was added and the solution was vortexed, then centrifuged at maximum speed in a microcentrifuge at room temperature for 20 minutes. The supernatant was removed and the pellet was washed with 0.5 ml of 80% ethanol. The solution was centrifuged at maximum speed for 5 minutes at room temperature. The 80% ethanol was poured off, and the 80% ethanol wash repeated. The pellet was air dried for approximately 15 minutes, then resuspended in 16 µl of nuclease-free water.

In Vitro Transcription: Using an Ambion T7 Megascript kit (Ambion, Austin, TX 78744-1832, USA), the manufacturer's instructions were followed to create a 40 µl reaction (*i.e.*, the 20 µl

standard reaction was doubled and incubated at 37°C for 4-5 hours). The reaction was assembled at room temperature using 16µl of template double-stranded DNA, to avoid the precipitation of spermidine, which can occur if done on ice. Four µl of 10x reaction buffer were added, then 4 µl of ATP solution (75mM T7), 4 µl of CTP solution (75mM T7), 4 µl of GTP solution (75mM T7), 4 µl of UTP solution (75mM T7), 4 µl of Enzyme Mix, and the reaction was incubated at 37°C for 5-6 hours. Sixty µl nuclease-free water was added to bring the total volume up to 100 µl. The RNA was "cleaned" using the "RNA clean-up" protocol provided in the Qiagen RNeasy Mini Handbook, May 1999, pp. 48-49, Qiagen, Valencia CA. RNA was eluted with 30 µl of nuclease-free water, and the optical density ratio was measured (the sample should have an optical density of greater than 1.8 when measured at 260/280 nanometers). The expected yield from this preparation was ten times the starting amount of total RNA, and the RNA was then ready for use in generating probe for microarrays using total RNA (see below).

Second Round Amplifications: 0.5-1.0 µg of amplified RNA were resuspended in 11 µl ultrapure water.

First Strand Synthesis: One µl Random hexamer (1 mg/ml) was added and the reaction was incubated at 70°C for 10 minutes, then chilled on ice, then allowed to equilibrate at room temperature for 10 minutes. Four µl 5X First strand cDNA buffer, 2 µl 0.1M DTT, 2 µl 10mM dNTP mix, 1µl RNasin were added, and the reaction was mixed incubated at 42°C for 2 minutes. Two µl Superscript II were added, and the reaction was mixed well and incubated at 42°C for 1 hour. One µl RNAse H was added, and the reaction was incubated at 37°C for 20 minutes, then heated to 95°C for 2 minutes to quell the reaction, then chilled on ice.

Second Strand Synthesis: One µl T7-oligodT primer (0.5 mg/ml) was added, and the reaction was incubated at 70°C for 5 minutes and at 42°C for 10 minutes. Then, 91 µl DEPC treated H_2O were added, then 30 µl Second strand buffer, 3 µl10 mM dNTP mix, 4 µl DNA Polymerase I (10U/ µl), 1 µl DNA Ligase (10U/µl), and 1 µl RNAse H (2U/µl) to a final volume of 150 µl. The tube was tapped gently to mix, then briefly centrifuged. The reaction was incubated at 16°C for two hours, then 2 µl (10U) T4 DNA Polymerase were added, and the reaction was cooled for 5 minutes at 16°C. The reaction was stopped with 10 µl of 0.5 M EDTA, then 10 µl of 1M NaOH were added. The reaction was incubated at 65°C for 10 minutes, then neutralized by addition of a solution with 25 µl Tris-HCl (pH=7.5). The protocol for "Clean Up of Double Stranded cDNA" and "In Vitro Transcription" was followed to generate cDNA for use in preparation of the probe for microarray hybridization.

Preparation of Probe and Microarray Hybridization Using Amplified RNA: To prepare the probe, reverse transcription labeling reaction mixes were created for each probe containing the component Random Primer (InVitrogen, Carlsbad, California 92008, USA). Three µg/µl in 2µl were added, then 5-6 µg amplified RNA. The reaction was brought to a final volume of 17 µl with water, then incubated at room temp for 10 minutes. To each probed, 5X first strand buffer in 8 µl, 20X lowT-dNTP mix in 2 µl, 0.1 M DTT in 4 µl, RNAsin in 1 µl, Cy-3 or Cy-5 dUTP (NEN Life Science, Boston, MA 02118-2512 USA) in 4 µl, and SuperScriptII (GIBCO-BRL, InVitrogen Corporation,

Carlsbad, California 92008 USA) enzyme in 2 µl was added. The reaction was incubated at 42°C for 60 minutes, then 5 µl of 500 mM EDTA and 10 µl of 1M NaOH were added. The reaction was incubated at 65°C for 15 minutes to hydrolyze residual RNA, then cooled to room temperature. Twenty-five 25 µl of 1 M Tris-HCl (pH7.5) was added to neutralize pH.

Probe Cleanup: 500 µl of 1X TE were added to a Microcon-YM30 column and the column was spun at 13,000 rpm for 5-6 minutes to wash the column. Membrane integrity was checked by looking into the top insert, to confirm that a thin film of TE (~50 µl) covered the membrane. 400µl 1X TE was added to each of the sample tubes and all contents were transferred to the washed Microcon-YM30 column (Amicon, Millipore Corp., Bedford, Massachusetts). The column was spun at 13,000 rpm for 5-6 minutes until approximately 50 µl was left on the membrane. The column was checked for dye crystals along the edge of the column membrane, which indicated that the probe was likely to be good. 450µl 1X TE were added to the column and the column was spun down to ~50µl as above. The presence of crystals was confirmed. The Cy-3 labeled probe was placed into a clean tube, and the column was spun at 14,000 rpm for 1 minute to elute the probe. The Cy-3 labeled probe was added to the Cy-5 labeled probe in the column, and approximately 450 µl 1X TE was added to the column. The column was spun at 13,000 rpm until approximately 13-14 µl of combined probe remained on the membrane, which was checked with a pipette. The combined probe was inverted into a clean tube, and spun at 14,000 for 1 minute to elute. The probe (14 µl) was transferred into a clean Eppendorf tube and stored at 4 °C until used in the hybridization reaction.

Probe Hybridization: Twenty µl of water were added to each humidifying well in the Hybridization Chamber (to maintain humidity). Then, 40 µl of prehybridization buffer (5X SSC, 0.1% SDS, 1% BSA (Sigma) warmed to 42°C) were placed in the center of the slide and the cover slip was placed on the slide, taking care to prevent bubble accumulation beneath the slip. The margin clamps on the Hybridization Chamber were firmly attached, and the chamber was incubated at 42°C for least 1 hour. The slide was washed in distilled water for 2 minutes, followed by isopropanol for 2 minutes. The slide was dried in a centrifuge (5804R, Eppendorf) at 705 rpm (~70x *g*) for 4 minutes, then prepared for hybridization as discussed above. The slide was hybridized within 1 hour of the prehybridization step. Two 2 µl COT1-DNA (Hoffman La Roche, Nutley, New Jersey 07110 USA) (1 µg/µl), 2 µl polyA (Sigma) (8-10 µg/µl), and 2 µl yeast tRNA (Sigma, Ronkonkoma, NY 11779 USA) (4 µg/µl) were mixed with the probe. Then, the probe was denatured for 1 minute at 100°C, placed briefly on ice to cool the reaction, and spun down in a centrifuge. Twenty µl of 2X hybridization buffer (50% formamide, 10X SSC, 0.2% SDS, warmed to 42°C) were added to the denatured probe, mixed well (taking care to minimize bubble formation) and kept at 42°C until ready to spot on the slide. The hybridization chamber was prepared as in the prehybridization step with 20µl of distilled water in each well. The slides were placed face-up in the chambers, and the probe was hybridized with the slide for 14-16 hrs at 42°C.

Slide Washes: The margin clamps on the Hybridization Chamber were carefully removed to prevent water from seeping in and contaminating the array. The slide was removed from the chamber, held with forceps and the cover slip allowed to fall off into the solution containing 2X SSC,

0.1% SDS. The slide was washed for 4 minutes in 1X SSC, 0.1% SDS, for 4 minutes in 0.2X SSC, and for 1 minute in 0.05X SSC. The slide was spun dry in a centrifuge at 705 rpm (approximately 70x *g*) for 4 minutes. If water droplets were seen on the slide, it was spin again for another 4 minutes. Exposure to light was minimized by placing the dried slides in a slide box until ready for scanning.

Statistical Analysis: The logarithmic expression ratios for the spots on each array were normalized by subtracting the median logarithmic ratio for the same array. Data were filtered to exclude spots with size less than 25 μm, intensity less than 2 times background or less than 300 units in both red and green channels, and any poor-quality or missing spots. In addition, any features found to be missing or flagged in greater than 10% of the arrays were not included in the analysis. Application of these filters resulted in the inclusion of 6,445 of the total 7,651 features in subsequent analyses. Statistical comparison between tumors groups was performed using the "BRB Array Tools" software (developed by Dr. Richard Simon and Amy Peng, Biometrics Research Branch, Division of Cancer Treatment and Diagnosis, NCI, USA). A modified F-test is run on each gene's log-ratio values, and the significance of that gene is determined with $P < 0.0001$ considered significant. This stringent *P* value is selected in lieu of the Bonferroni correction for multiple comparisons, which was deemed excessively restrictive (Figure 1) (see Bland and Altman, *B.M.J.*; 310: 170, 1995).

Semi-quantitative PCR: Five samples from each tumor group (*BRCA*-1, *BRCA*-2, sporadic) were selected at random. For each sample, 3.5 μg of total RNA was reverse-transcribed using oligo dT primers and 400 units of Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA) in the presence of all four deoxyribonucleoside 5'-triphosphates (each at 10 mM) (InVitrogen, Carlsbad, CA) and 40 units of RNAse inhibitor (Promega, Madison, WI). Reverse transcription was performed in a total reaction volume of 40 μl, of which 1 μl was subsequently used for each PCR reaction. Preliminary experiments were performed to identify optimal cycle number for each gene. Thirty cycles was found to be optimal amplification for all amplified RNAs except for *HLA-DRB1* and *CD74*, which were amplified for 26 cycles. Polymerase chain reaction was performed using the GeneAmp PCR kit (PE Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. Representative gene specific primer sequences are shown in Table 3: